5-1997

The Relationship Between Graduate Students' Education in Research Ethics and Their Attitudes Toward Research Misconduct

Perry Sailor
Utah State University

THE RELATIONSHIP BETWEEN GRADUATE STUDENTS' EDUCATION IN

RESEARCH ETHICS AND THEIR ATTITUDES TOWARD

RESEARCH MISCONDUCT

by

Perry Sailor

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Psychology
(Research and Evaluation Methodology)

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1997

ABSTRACT

The Relationship Between Graduate Students' Education in

Research Ethics and Their Attitudes Toward

Research Misconduct

by

Perry Sailor, Doctor of Philosophy

Utah State University, 1997

Major Professor: Dr. Blaine R. Worthen
Department: Psychology

A mail survey of a nationwide sample of department heads in university departments of mechanical engineering, physiology, and psychology was conducted, in order to determine what these departments were doing to educate their Ph.D. students in research ethics. Department heads were also asked to supply names of the Ph.D. students in their departments. Based on the survey responses, departments within each discipline were then divided into those placing a relatively high versus low emphasis on teaching research ethics. Random samples of students in each emphasis category for each discipline were then surveyed and asked to rate the seriousness of 44 different hypothetical acts of misconduct, to determine if students from departments placing relatively higher emphasis on research ethics education had stricter standards than those from departments placing relatively lower emphasis on research ethics education. The two major findings of the study were (a) the majority of departments in physiology and psychology require some form of formal education in research ethics of their Ph.D. students, but only a very small percentage of mechanical engineering departments require such training; (b) the present study found no evidence that education of Ph.D. students in research ethics has any effect on the strictness of their stated ethical standards.

(143 pages)

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Blaine R. Worthen, for his suggestion of the topic for this dissertation and for his invaluable support and guidance throughout the dissertation process and throughout my entire graduate school career.

I would also like to express my appreciation and gratitude to the other members of my committee, Drs. Byron Burnham, Xitao Fan, G. Jay Gogue, and Lani Van Dusen, for their helpful suggestions for improving both the study itself and this dissertation.

My appreciation and gratitude also go to each of the busy department heads and graduate students who took time to complete their respective surveys. Also, my sincere thanks to some companions who helped make at least somewhat enjoyable the sometimes tedious tasks of survey preparation, envelope addressing and stuffing, data analyzing, and writing, including among others Antonio Vivaldi, Johann Sebastian Bach, Bob Dylan, Miles Davis, Duke Ellington, and Van Morrison.

Finally, I would like to give special thanks for their love and support to my late parents, Raymond J. (Al) and Wilma Sailor, to Lori Roggman, to my children, Rae and Mark, and to the good and special friends who helped keep my spirits up and get me through this whole process--you know who you are.

Perry Sailor

CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The entire scientific enterprise rests on an assumption of honest, ethical behavior by scientists (Aronson, 1975; Freedman, 1992; Medawar, 1990b; National Academy of Sciences, 1989; Relman, 1989). Indeed, the ethical norms of science are so strongly held that when serious ethical breaches come to light, the response by the scientific community and the public at large is one of outrage. Some commentators have labeled these reactions as "out of proportion to their immediately pragmatic rather than symbolic consequences" (Zuckerman, 1977, p. 89), with such terms as "heinous" (Medawar, 1990a, p. 91), and "a quality of desecration" (Luria, 1975, p. 15) used to describe them. As one scientist has put it, "...[F]alsifying evidence is just about the worst sin that a scientist can commit since such actions threaten to destroy the very heart of the scientific system" (Aronson, 1975, p.115).

The ethical codes of science are rooted in what sociologists of science (e.g., Merton 1973; Zuckerman, 1977) have termed the "ethos" or normative structure of science. Most types of misconduct can be classified as violations of one or more of these norms, enumerated by Merton (1973) as follows: (a) universalism (truth claims to be subjected to preestablished impersonal criteria, unrelated to personal characteristics of their authors); (b) communism (findings of science are a product of social collaboration and belong to the community at large); (c) organized skepticism (truth claims must be subjected to the detached scrutiny of the scientific community); and (d) disinterestedness (unclearly defined by Merton, but explicated by Zuckerman [1977] as disinterested activity intended to extend scientific knowledge).[1]

---

[1]Storer (1973) believes disinterestedness is perhaps best understood as similar to the 17th-century Protestant concept of stewardship, or calling.

Other authors have added to Merton's norms those of individualism (following one's own judgment), rationality, and emotional neutrality (Barber, 1952); and honesty, objectivity, and unselfish engagement (Cournand & Zuckerman, 1970; Zuckerman, 1977). In contrast to Merton's terms, the meanings of these additional norms are manifest and need not be explained here.

With much scientific research, especially in the life sciences, becoming increasingly "high stakes" in monetary terms, and with many Ph.D. programs at state-supported universities coming under increasing legislative pressure to move students quickly through their programs (G. Jay Gogue, personal communication, November 27, 1996)--pressure that is transmitted in turn to the students themselves, in the form of stringent time limits for successful program completion--research ethics is an important issue now and likely to become even more important in the future. Indeed, several high-profile cases of ethical misconduct have occurred in recent years (discussed by Bell, 1992; Braunwald, 1992; Broad & Wade, 1982; Kevles, 1996; Kohn, 1986; Medawar, 1990a, 1990b; Miller, 1992b; Poling, 1992; Rensberger, 1977; Szilagyi, 1984). Along with such phenomena as a Food and Drug Administration report of evidence of "significant misconduct" in 11% of 1,758 "essentially random" audits conducted over 10 years (DuBois, 1989, p. 606), these cases have brought many calls for various reforms, even including a massive overhaul of the entire funding-agency system within which so much research is conducted (Bell, 1992; Broad & Wade, 1982).

Among the less drastic reforms suggested, one of the most frequent may be expressed in terms of the familiar bit of folk wisdom "an ounce of prevention is worth a pound of cure." To put it in more academic terms, it has often been speculated that misconduct can be prevented by teaching research ethics at the graduate school level (Hilgartner, 1990; Institute of Medicine, 1989; Kromrey, 1993; Mishkin, 1988; National

Academy of Sciences, National Academy of Engineering, & Institute of Medicine, 1992; Relman, 1989; Steneck, 1994).

However, this proposed remedy (and indeed the entire discussion of research misconduct) has taken place in a virtual vacuum of empirical knowledge. Although much has recently been written about individual universities' attempts to address ethics education (e. g., Steneck, 1993; 1994), an extensive search by the present author turned up little knowledge about (a) the extent to which research ethics is formally or informally taught in graduate science programs, (b) the relative prevalence of different means of teaching research ethics, (c) how instruction in ethics may differ across disciplines, or (d) whether such instruction, where it exists, has any impact on students' actual research conduct, or even on their thinking about the conduct of research. Given the importance of ethical research conduct to the scientific enterprise, and given the widespread belief that instructing graduate students in research ethics will reduce the incidence of misconduct, it is clear that such empirical knowledge is vital.

A properly done survey of department heads and graduate students in several science disciplines would provide needed information concerning the present state of ethics instruction in graduate training in the sciences. Such a study would also provide information concerning whether the inclusion of formal instruction in research ethics in Ph.D. training is correlated with how students feel about the seriousness of various research practices identified as possibly unethical. This study is proposed to provide such information.

Research Objectives

The general purposes of the proposed study are to determine the present state of affairs regarding ethics education in graduate departments in psychology, biology (physiology), and mechanical engineering, and to see if ethics education of graduate students is related to their standards of research ethics. The specific objectives are (a) to determine the relative prevalence of various alternative methods of providing research ethics education: day-to-day informal faculty contact (i.e., no formal research ethics education), formal classes devoted to ethics education, formal classes mostly devoted to other content but including a research ethics component, informal seminars or discussion groups on research ethics, and written handbooks and other policy statements; (b) to determine the relative frequency with which research ethics education is provided at different administrative levels (i.e., department, college, graduate school, university); (c) to determine the relative frequency of required versus elective research ethics education; (d) to determine if students in departments that have formal education in research ethics have more rigorous ethical standards than those in departments that do not provide ethics education; (e) to determine if, among departments that do provide research ethics education, there are differences in standards of rigor related to methods of ethics education employed; and (f) to assess how graduate students rate the relative seriousness of different kinds of research misconduct.

Research Questions

The research questions to be answered by this study are as follows:

1. What are university departments, colleges, and central administrations doing to teach research ethics to graduate students? That is, what is the relative prevalence of such methods as formal classes, informal seminars or discussion groups, lectures, and written handbooks or other written policy statements? Where provided, is research ethics education handled at the department level, the college level, the university level,

or some other? In how many departments that provide research ethics education is it required, rather than elective? How many departments do not provide research ethics education at all?

2. Are the relative frequencies of different methods of research ethics education different in different disciplines? Do they differ according to the quality rating or size of the department?

3. What is the difference between students in departments with formal research ethics education and those in departments without formal research ethics education in terms of how they judge the seriousness of various kinds of possible misconduct?

4. Among students in departments with formal research ethics education, what is the difference between students in departments in which research ethics education is required and students in departments in which it is not required, in terms of how they judge the seriousness of various kinds of possible misconduct?

CHAPTER II

REVIEW OF LITERATURE

Defining Research Misconduct

Theoretical Basis for Defining Misconduct: The Norms of Science

Whatever specific acts are defined as some kind of research misconduct, and however a given act may fit into some category of misconduct, most, if not all, research misconduct can be characterized as a violation of the norms or ethos of science. Accordingly, this section is a discussion of these norms.

Merton (1942/1973), in a seminal theoretical article, labeled the moral or social norms of science as universalism, communism, disinterestedness, and organized skepticism. He called these norms "four sets of institutional imperatives...[that] are taken to comprise the ethos of modern science" (p. 270). Their meanings may be understood as follows: Universalism imposes the requirement that truth-claims are to be subjected to preestablished impersonal criteria, having nothing to do with the personal or social attributes of their authors. The norm of communism--or as Barber (1952) preferred, communality[2]--holds that the findings of science are a product of social collaboration and are assigned to the community, with intellectual "property" being converted to the coin of peer recognition, rather than its right of use being retained by its creator.

Disinterestedness is not, Merton (1942/1973) said, a characteristic of scientists themselves, but rather a characteristic of the institution of science that may (or may not)

[2]Barber's usage was probably a product of the 1952 version of political correctness, and the present investigator is tempted to stick with Merton's terminology. However, in current usage, the term "communality" is much closer to Merton's meaning than is "communism," which has come to be associated with a particular sociopolitical system. Accordingly, "communality" is used in the remainder of this proposal.

be internalized by scientists. However, "once the institution enjoins disinterested activity, it is to the interests of scientists to conform on pain of sanctions and, insofar as the norm has been internalized, on pain of psychological conflict" (p. 276). Merton attributed science's "virtual absence of fraud" (p. 276) to this norm, which is demanded by the "public and testable character of science," and which subjects results to the "exacting scrutiny of fellow experts." (p. 276). According to Merton, then, scientists aren't any better or more moral than anyone else, it is just that the nature of the social structure of science makes fraud futile and self-defeating.

The fourth of Merton's (1942/1973) norms, organized skepticism, requires that scientific claims be subjected to "the detached scrutiny of beliefs in terms of empirical and logical criteria" (p. 277), is interrelated with the other three elements, and is part of both methodology and of science as an institution.

Taken together, these four norms cover most of the serious forms of scientific misconduct mentioned in the literature. As Merton's former student Zuckerman (1977) summed up, "Violations of these moral norms of science include ad hominem attacks (antiuniversalism); the contriving of forged or otherwise fraudulent evidence (antidisinterestedness); plagiarism and secrecy (anticommunality); dogmatism and shoddy work (antiorganized skepticism), and a range of other actions variously proscribed by these norms" (p. 89).

Many authors since Merton have used his delineation of scientific norms to organize their own discussions of scientific ethics, but have elaborated upon them or expanded them. For example, Barber (1952), added "individualism" (meaning "antiauthoritarianism" or following one's own judgment), "rationality," and "emotional neutrality" to Merton's norms. Merton himself (1957) cited Barber's list in a reference to "other values of science" (p. 646), and added "humility" to his own list.

Cournand and Zuckerman (1970) termed the scientific norms "honesty," "objectivity," "tolerance," "doubt of certitude," and "unselfish engagement." Mohr's (1979) list of normative rules is put in more specific terms than Merton's, but each can be subsumed under them: "Be honest; never manipulate data; be precise; be fair with regard to priority; be without bias with regard to data and ideas of your rival; do not make compromises in trying to solve a problem" (p. 48).

The reader may have observed that researchers, in their everyday activities, are often anything but disinterested or communal in their outlook. Indeed, other authors who have followed Merton have expanded on and refined his concepts by demonstrating (or, in some cases, theorizing) how different situations may affect the applicability of the norms. Ben-David (1991), for example, believed scientists do subscribe to and behave in accordance with Merton's norms when it comes to allocating scientific rewards or sitting in judgment of peers. But the norms do not guide scientists' behaviors in the more disorderly, early, "solution of problems" stage of research on a given issue. To use a legal metaphor, this stage involves the processes preceding judgment--sifting evidence, determining the relevant issues, presenting the case to judge and jury. In this stage, scientists are litigants or advocates rather than judges. They become judges--or at least submit to the opinions of judges--after all the evidence is in and they have made their best case. This is the point at which Merton's norms apply to science.

Zuckerman (1977), like Ben-David, pointed out that different norms are relevant at different stages of the scientific process. She called the early stage the "private phase," (p. 124), and believed that disinterestedness or skepticism at this stage is inappropriate.

Goodstein (1995) has recently written that the view of the scientist as disinterested observer, which he traced back to Francis Bacon, has been discredited by philosophers and historians of science, but still exists in the public mind as the "Myth of the Noble Scientist," who is supposedly "more upright and virtuous than ordinary mortals, impervious to the baser human drives, such as personal ambition, and, of course, incapable of misbehaving in even the smallest way" (p. 31).

While at first glance this may seem to be a refutation of Merton's disinterestedness norm, it is the present author's opinion that there is actually no conflict between Merton and Goodstein here. Recall that Merton (1942/1973) freely admitted that scientists are not inherently more moral than others, and that disinterestedness is a characteristic required by the institution of science, and thus constrains behavior whether or not it is internalized by the individual scientist.

While the present investigator has not found any published denial that some kind of ethos of science exists, an extensive search of the literature found that few researchers have put Merton's specific formulation to any kind of explicit empirical test. One exception is a study by Toren (1980) that examined the applicability of traditional ethos to modern Israeli scientists, all of whom had emigrated from either the U.S.S.R. or the U.S. within the past few years. She did an attitude survey on Merton's norms, which she called the "traditional ethos" of science. Toren found that the majority of scientists did not subscribe to the norms of emotional neutrality (from Barber, 1952), organized skepticism, and communality; that disinterestedness was accepted by about half; and that only universalism was adhered to by as many as two thirds. However, Toren's survey questions were mostly couched in terms of the early stages of research, so her results support the notion of different norms applying to different stages, rather than that the norms simply do not apply.

Mitroff (1974) tested Merton's norms explicitly in terms of their situational applicability. Based on a case study of Apollo moon scientists, Mitroff concluded that conventional scientific norms seemed to be dominant for well-defined problems, but that "counternorms" were dominant for ill-defined problems. The counternorms include "emotional commitment," "particularism," (the opposite of universalism, i.e., status matters), "solitariness," "interestedness," and "organized dogmatism." Mitroff's findings, like Toren's (1980), seem to support the speculations of Ben-David (1991) and Zuckerman (1977). All types of misconduct in the domain of the present study are related to the later phase of research, when problems are well defined and the research is being presented for the public judgment of peers and the larger public. Accordingly, the traditional norms would be expected to apply, and violations of them would in fact constitute misconduct.

Defining Misconduct in Practice

In practice, there is no single agreed-upon definition of research misconduct. Authors who have written about misconduct typically either have provided their own definition or else left it undefined and assumed the reader knows what they meant. The latter approach is probably adequate when the conduct in question is such obviously fraudulent behavior as making up data or out-and-out plagiarism. However, there are many acts of misconduct that fall short of fraud, some of which may not even be universally considered wrong. For example, Rosenthal (1994) recently asserted that unethical actions extend to such conduct as using a poor research design, failing to fully exploit data (meaning failure to break the data down in originally unintended ways, or to look for unpredicted relationships, things that most authorities warn against), and doing a traditional narrative literature review in a quantitative field in which meta-analytic methods could and should be used. All of these, Rosenthal maintained, waste precious resources, whether they be data or simply time and effort.

Definitions from Individual Authors

In discussing research misconduct, Thelen and DiLorenzo (1992) restricted the term research fraud to acts involving intent to deceive, intentional misrepresentation, gross mismanagement, and possible punitive consequences from outside academia. In discussing the larger realm of ethical violations they preferred the term "intellectual dishonesty," which covers a wider range of behavior, including other types of misconduct and misrepresentation as well as fraud. Intellectual dishonesty does not require the intent to deceive, but rather includes such forms of misconduct as carelessness and even some conduct that some might not consider unethical, such as repeated publication of the same data. Engler, Covell, Friedman, Kitcher, and Peters (1987) gave other examples of borderline ethics, such as "to forget an 'anomalous' result, to fail to scrutinize data in a collaborative project, to leave the collection of data to a fast-working colleague known to be 'sloppy,' or to accept gift coauthorship" (Engler et al., 1987, p. 1384).

In a similar fashion to Thelen and DiLorenzo (1992), Szilagy (1984), Schaffner (1992), and Zuckerman (1977) have also made some distinction in degree of culpability. Szilagy distinguished "deceit" from "fraud," and considered the latter more serious. Schaffner called fraud the "intention to deceive the scientific community about the nature of scientific results..." (p. 18), and distinguished it from other breaches of norms. Zuckerman likewise divided deviations from moral norms into "fraud," "other violations of moral norms," and "deviations from etiquette." "Fraud" can involve either fabrication, fudging, or suppression of data. "Other violations" include such offenses as plagiarism, suppression of radically new truth claims, and secrecy. "Deviations from etiquette" include eponymization, underacknowledgment of collaborators, ad hominem attacks, and publicity seeking. Zuckerman also considered "disreputable error" (that is, those resulting from neglect or violation of proper methodology) a kind of misconduct, terming it the "counterpart to legal negligence" (p. 110)--except that science, unlike the law, considers willful negligence and inadvertent negligence to be equally serious offenses.

Formal Definitions from Organizations

Besides individual authors, other sources of definitions of research misconduct are such research sponsors as the National Institutes of Health (NIH) and the National Science Foundation (NSF). Spurred by several high-profile fraud cases and the resulting increased Congressional attention in the early 1980s, these organizations scrambled to produce formal procedures for dealing with misconduct, in which a necessary step was defining it (Price, 1994).

In 1986, the NIH defined misconduct as "(1) serious deviation, such as fabrication, falsification, or plagiarism, from accepted practices in carrying out research or in reporting the results of research; or (2) material failure to comply with Federal requirements affecting specific aspects of the conduct of research, for example, the protection of human subjects and the welfare of laboratory animals" (Public Health Service, 1986, quoted in Price, 1994, p. 287). The following year, NSF adopted a similar definition. In 1989 (Department of Health and Human Services, 1989), the Public Health Service (PHS) added "or other practices that seriously deviate from those that are commonly accepted within the scientific community" (Price, 1994, p. 288). This addition was very controversial, with some commentators fearing that it was overly vague and could even stifle potentially creative or innovative challenges to established methods or beliefs, and others criticizing that it did not distinguish serious from minor infractions. The PHS also added that misconduct "does not include honest error or honest differences in interpretations or judgments of data" (Price, 1994, p. 288).

The National Academy of Sciences (NAS et al., 1992) next weighed in with its definition. The NAS defined research misconduct as "fabrication, falsification, or plagiarism, in proposing, performing or reporting research" (p. 5). The NAS also excluded errors and differences of opinion from its definition, and rejected the PHS's vague "other practices that seriously deviate from those that are commonly accepted..." phrase. However, the NAS defined a second, less serious category of misconduct, "Questionable Research Practices," which it defined as "actions that violate traditional values of the research enterprise and that may be detrimental to the research process" (p. 5). They gave as examples such conduct as failing to retain research data for a reasonable period, maintaining inadequate records, conferring or requesting authorship on the basis of service or contributions not significantly related to the research reported, refusing to give peers access to data, using inappropriate statistical methods to enhance the significance of the findings, inadequately supervising subordinates or exploiting them, and misrepresenting speculations as fact or releasing preliminary results to the public media in advance of peer review. In 1991, the NSF added retaliation against whistleblowers to its definition of misconduct. Price's conclusion: The status of these definitions is fluid and likely to remain so for some time (Price, 1994).

Specific Actions That May Be Considered Unethical

What kinds of actions can be considered "research misconduct"? Up to this point, the present author has treated misconduct as a fairly abstract concept, even while providing definitions given by various authors. Before going further into the discussion of the literature, some more specific examples of misconduct seem in order. Table 1 contains a list of 49 possible acts of misconduct, covering the gamut of offenses, that

Table 1

Types of Misconduct Cited in the Literature

Type of Misconduct	Authors
Misconduct Related to Methodology	
1. Carelessness or bias in conducting or recording experiments.	Barber, 1976; Relman, 1989
2. Fabrication of data.	Anderson, Louis, & Earle, 1994; Bell, 1992; Ghiselin, 1989; Kalichman & Friedman, 1992; Mahoney, 1976; Medawar, 1976/1990a, 1983/1990b; Merton, 1957; Schaffner, 1992; St. James-Roberts, 1976a, 1976b; Swazey, Anderson, & Louis, 1993; Tangney, 1987; Zuckerman, 1977
3. Failure to disclose weaknesses in data or research design.	Bailar, 1986
4. Failure to make raw data available for re-analysis.	American Statistical Association, 1983; Craig & Reese, 1973; Robinson & Moulton, 1985; Wolins, 1962
5. Neglect or violation of methodological concerns and procedural precautions, (e.g., loosely following experimental protocol).	Barber, 1976; Cohen & Ciocca, 1992; Miers, 1985; Zuckerman, 1977
6. Rigging experiments.	St. James-Roberts, 1976a, 1976b
7. Using a poor research design.	Rosenthal, 1994
Misconduct Related to Data Analysis and Interpretation	
1. Cutting data for analysis in originally unintended ways.	Barber, 1976; Mahoney, 1976
2. Double checking only negative results.	Barber, 1976; Engler, Covell, Friedman, Kitcher, & Peters, 1987
3. Fudging (i.e., selective deletion) or suppression	Hagstrom, 1974; Kromrey, 1993; of data points. Mahoney, 1976; Schaffner, 1992; St. James-Roberts, 1976b ["data massage"]; Zuckerman, 1977

(table continues)

Type of Misconduct	Authors
4. Honest error (e.g., inadvertent errors in recording or reporting data, or incorrect but honestly reached conclusions about the results).	Price, 1994; Rosenthal, 1994
5. Incompetent data analysis.	Altman, 1980
6. Overlooking others' use of flawed data, questionable interpretations, or other research transgressions.	Anderson et al., 1994; Engler et al., 1987, Hilgartner, 1990
7. Reporting statistical significance while ignoring effect size.	Barber, 1976; Kromrey, 1993
8. Selection and manipulation of results.	Blakely, Poling, & Cross, 1986
9. Statistical testing of post hoc hypotheses.	Bailar, 1986; Barber, 1976; Kromrey, 1993
<u>Misconduct Related to Publication</u>	
1. Failure to disclose weaknesses in data or research design.	Bailar, 1986
2. Failure to present results that contradict one's previous research.	Anderson et al., 1994; Rosenthal, 1994
3. Failure to publish negative results.	Chalmers, 1990; Ghiselin, 1989; Mahoney, 1976; Merton, 1957; Rosenthal, 1994; Schaffner, 1992; Zuckerman, 1977
4. Failure to publish until follow-up work is complete.	Hagstrom, 1965; Merton, 1957; Zuckerman, 1977 [secrecy]
5. Giving only a cursory review to a paper submitted for publication, if it supports one's own theory.	Hilgartner, 1990
6. Incomplete authorship.	Huth, 1986; Schaffner, 1992
7. Incomplete documentation of work.	American Statistical Association, 1983
8. Intentional efforts to communicate false or misleading findings.	Bobys, 1983
9. Misinterpretation of results.	Robinson & Moulton, 1985; St. James-Roberts, 1976b
10. Misrepresentation of another's work in a citation.	Ghiselin, 1989
11. Misrepresentation of data, research procedures, or data analysis.	Garfield, 1987; Mahoney, 1976; Mishkin, 1988; Rosenthal, 1994
12. Multiple papers from one study.	Huth, 1986

<u>(table continues)</u>

Type of Misconduct	Authors
13. Plagiarism.	Anderson et al., 1994; Ghiselin, 1989; Hagstrom, 1974; Kalichman & Friedman, 1992; LaFollette, 1992; Merton, 1957; Miller, 1992b; Schaffner, 1992; Swazey et al., 1993; Tangney, 1987; Zuckerman, 1977
14. Publication of same material repeatedly.	Huth 1986
15. Selective reporting of findings.	Bailar, 1986; Barber, 1976; Birch, 1990; Engler et al., 1987; Kromrey, 1993; Mahoney, 1976; Medawar, 1983/1990; Merton, 1957; Miers, 1985
16. Underacknowledgment of intellectual predecessors, rivals, colleagues.	Birch, 1990; Garfield, 1980; Merton, 1957 ["aggressive self-assertion"]; Schaffner, 1992; Zuckerman, 1977
17. Unjustifiable authorship.	Anderson et al., 1994; Cohen & Ciocca, 1992; Engler et al., 1987; Garfield, 1980; Ghiselin, 1989; Huth, 1986; Kalichman & Friedman, 1992; Swazey et al., 1993
Misconduct Related to Personal Integrity	
1. Ad hominem attacks.	Zuckerman, 1977
2. Applying for funding to support work already done.	Ghiselin, 1989
3. Failure to adequately inform human subjects.	Cohen & Ciocca, 1992; Robinson & Moulton, 1985
4. False claims or commitments made in grant proposals.	Harrobin, 1989
5. Misrepresentation of publication status of an article.	LaFollette, 1992
6. Mistreatment of human or animal subjects.	Miers, 1985
7. Retaliation against whistle-blowers.	Price, 1994
8. Reviewing others' work unfairly (e.g., to sabotage a rival).	Ghiselin, 1989; Zuckerman, 1977
9. Slanderous charges of plagiarism.	Merton, 1957
10. Stolen ideas.	Hagstrom, 1965; Mahoney, 1976; Merton, 1957; Robinson & Moulton, 1985; Schaffner, 1992; Steneck, 1984
11. Trying to get by on the work of others.	Hagstrom, 1965

(table continues)

Type of Misconduct	Authors
12. Ultimate outcome of research having bad consequences, for example, work with genetically altered microorganisms unleashing a deadly plague, or behavioral genetics research leading to racial discrimination, or research having military applications).	Robinson & Moulton, 1985
13. Using one's position to exploit or manipulate others.	Hagstrom, 1965; Rosenthal, 1994
14. Using university resources for outside consulting work or other inappropriate personal purposes.	Bell, 1992; Swazey et al., 1993
15. Violation of federal, state, or institutional rules.	Mishkin, 1988
16. Violation of privacy or confidentiality norms regarding subjects.	American Statistical Association, 1983

the present author has extracted from the literature on research misconduct. (The format of Table 1, and a few of the elements, are loosely based on a much smaller table in Thelen & DiLorenzo, 1992.)

The phrase "research misconduct" probably evokes in most readers' minds such acts as making up data or outright plagiarism, and it is true that such "high crimes" have dominated the discussions in the literature. However, as Table 1 shows, the types of misconduct discussed in the literature cover a wide range of activities, many of which are less serious than outright fraud and some of which may not even be universally considered to constitute misconduct. While they may seem relatively less serious, these types of actions are nevertheless worthy of discussion and study for two reasons. First, it has often been asserted (Bailar, 1986; Birch, 1990; Engler et al., 1987; Kromrey, 1993; Merton, 1957), and seems intuitively true, that the less serious transgressions happen more frequently than outright fraud or plagiarism. Second, if this is true, the cumulative effect of the smaller offenses may be more insidious than the relatively few (one presumes) cases of outright fraud. In other words, a number of researchers unjustifiably deleting pesky outlier data points in order to make a weak relationship appear strong may do as much damage as one researcher simply making up data.

Reasons Why Scientists Commit Unethical Acts

Many authors have speculated as to why scientists would commit grievously unethical acts like fraud and plagiarism; however, to the present author's knowledge, there are no empirical studies to support these speculations, which, for the most part, tend to explain misconduct in terms of either (a) the individual (i.e., some people in any field of human endeavor are dishonest), (b) the institutional structure of modern scientific research (i.e., science is set up to encourage misconduct, or at least to fail to detect it), or (c) an interaction between the two (i.e., some characteristics of the system, for example, pressure to publish, pushes some vulnerable people over the edge into misconduct).

The structural view is represented by scientist/journalists Broad and Wade (1982), who in their very widely read and cited book, denied claims by "spokesmen for the conventional ideology of science....[that fraud is invariably]...the product of a deranged mind...[and is invariably discovered by]...self-policing mechanisms." (p. 7) Rather, they claimed that the system by which scientific research is funded and carried out is vulnerable to and creates pressures for fraud. "[The] roots of fraud lie in the barrel, not in the bad apples that occasionally roll into public view" (p. 9). Casti (1989) also blamed the system which pressures scientists to produce original work, and referred to "science's Faustian bargain with government funding agencies" (p. 52) as a causal factor in misconduct.

Hersen and Miller (1992) also presented a structural explanation, claiming that "...the heart of the problem [is] the commercialism of academia and its attendant issues" (p. 229, emphasis in original). By this, they meant the fault lies in academia as big business; the "publish or perish" rule and its subclause, "publish positive results or perish"; academic entrepreneurship; lack of careful mentorship; and an anti-intellectual climate. All of these, Hersen and Miller concluded, make cheating more likely. A "system" explanation for misconduct was also presented by Bobys (1983), who focused especially on competitive "publish or perish" policies.

Petersdorf (1986, 1989) blamed cheating in medical science on "pre-med syndrome," and asserted that the "culture in which we train our young [in medicine] promotes cheating" (1986, p. 252). He also blamed the large size of the medical science establishment and the intense competition for grants and tenure: "Medical science today is too competitive, too big, too entrepreneurial, and too much bent on winning" (1986, p. 253). Petersdorf's comments refer directly only to medical research, but, at least to some degree, they apply to all research based in universities and relying on competitive grant funding.

In contrast to the structural explanation, some authors have laid the blame for research misconduct squarely on the shoulders of the individual perpetrators. Kohn (1986), for example, explicitly disagreed with Broad and Wade (1982) about the roots of scientific fraud being systemic and endemic to the structure of modern science. On the contrary, Kohn believed all the pressure goes toward honesty, and that fraud in science is caused by a few dishonest individuals, of a kind likely to be found in any profession. He claimed, based on personal impressions gained through conversations with scientists, that most dishonest people are rooted out because of incidents that occur when they are graduate students or undergraduates, although the incidents are never reported (Kohn, 1986, p. 198-199).

The late president of the National Academy of Sciences, Philip Handler, argued that acts of fraud demonstrate "psychopathic behavior...minds which at least in this one regard may be considered deranged" (quoted in Woolf, 1981, p. 10). And Braunwald (1992) believed many instances of scientific fraud represent "a form of unconscious self-destructive behavior, with aggressive components directed also toward colleagues, supervisors, institutions, and society" (p. 76). Elsewhere, Braunwald (1987) added, "...Indeed, it mocks science as a whole" (p. 216).

Perhaps the most compelling explanations for misconduct are those that place exclusive blame neither on the individual nor the system, but rather on an interaction between them (Miller, 1992a). For example, many authors (Fox, 1989; Kubie, 1954; Mahoney, 1976) have remarked on the capriciousness of the peer review system and even of overall success in a scientific career. As Kubie (1954) pointed out, "[C]hance [is] a major factor in determining not what is discovered, but when and by whom" (p. 111). Kubie believed that this fact could cause young scientists to become "hardened, cynical, amoral, embittered, [and] disillusioned" (p. 112), which certainly sounds like someone prone to deviant behavior.

Other writers seeking an explanation for misconduct have emphasized the great rewards at stake, especially in certain lucrative areas like biomedicine. Faber (1974), in a letter to the editor of Science, said:

> We are naive to believe that dishonesty in research is unique and aberrant. The rewards are just too tempting: prestige, ego enhancement, promotion, and [financial benefits]....Not only are the rewards tempting but, while the process of socialization in graduate school may give credence to veracity, it nonetheless emphasizes success. The emphasis on scientific success creates a severe strain on the practicing researcher, who is torn between the norms established for the process of research and the penultimate rewards for success. Under these conditions deviance is likely to occur in any group, even among scientists. (p. 734)

Thelen and DiLorenzo (1992) enumerated the pressures that can accumulate on a researcher, but noted that ultimately these reduce to the amount of stress the person experiences. They believed that these factors--competition, "publish or perish," time pressure, and so forth--either alone or in concert with existing psychopathology, can make misconduct likely.

Knight (1984), a psychiatrist, also recognized that personal and structural factors may act jointly to cause misconduct. In Knight's view, nonpersonal factors that may contribute to misconduct include institutional pressures, competition for resources, public attitude (they demand results), shared interests with other scientists (leading to lack of scrutiny or vigilance), and a powerful cultural ideal of success. Personal factors contributing to misconduct may include stress and the consequent impairment of reasoning, lust for power and glory, quest for security through success, an ethics-neutral education system, and failure to reach the highest level of moral development.

Miller's (1992a) discussion of errors caused by unconscious self-deception includes several speculative instances of person-situation interactions resulting in misconduct. Possible factors discussed by Miller include (a) the actor/observer phenomenon, well known in social psychology (i.e., actors tend to make behavioral attributions that focus on factors inherent to the situation, while outside observers attribute behavior to stable personality characteristics of the actor); (b) psychodynamic/developmental factors (i.e., misconduct as displaced aggression against the self and others); (c) modeling of others' misconduct; and (d) cognitive dissonance (which a person may reduce through cognitive refocusing or rationalization).

The sociologists Bechtel and Pearson (1985) drew upon Merton's (1968) anomie theory in their effort to construct a "sociology of scientific deviance" (p. 245). Merton's theory holds that nonconforming behavior (such as misconduct) is a "symptom of dissociation between culturally prescribed aspirations and socially structured avenues for realizing these aspirations" (p. 188). "Any extreme emphasis upon achievement will attenuate conformity to...institutional norms" (p. 220). As explicated by Bechtel and Pearson, anomie theory places the motivation toward deviance on "frustrations encountered by those who are expected to achieve, even told to achieve, but lack legitimate resources to be successful" (p. 247). Bechtel and Pearson's (1985) application of anomie to scientific misconduct received indirect support from Barber, Lally, Makarushka, and Sullivan's (1973) finding that scientists who had been relative failures but who still strived to achieve success were more likely to do ethically suspect research with human subjects.

The Amount of Research Misconduct in Science

The literature that attempts to answer this question can be divided into two categories: answers based on speculation, and answers based on empirical studies.

Answers Based on Speculation

The general consensus in the literature on the amount of research misconduct in science is that it is unknown. As Miller and Hersen (1992) put it, "Although much anecdotal evidence exists, from an empirical perspective it is clear that research on base rates of scientific fraud and misconduct are lacking, and conclusions simply are at the speculative level" (pp. 4-5). Miller and Hersen's conclusions have been echoed by Bell (1992), Garfield (1987), Merton (1957), Thelen and DiLorenzo (1992), and Woolf (1981).

A few authors have been willing to venture informed opinions, without documentation, concerning the level of misconduct. Braunwald (1992), for example, asserted that "although the vast majority of scientists are honest, a small percentage cheat occasionally, usually under great pressure or when they feel the stakes are very high" (p. 75). And Koshland (1987), in an editorial in Science, claimed (without citing evidence) that 99.999% of scientific results are accurate and truthful. Earlier, Luria (1975) also claimed that serious misconduct is rare, although he added the qualification "or at least...rarely discovered" (p. 16). While also asserting that serious fraud is rare, Hagstrom (1965) added that a lot of fraud is simply ignored rather than challenged. Hagstrom's contention was supported by psychologist Leon Kamin's cynical view (Rensberger, 1977): "...[I]t's relatively easy to fake it and get away with it, particularly if it matches everybody's preconceptions....Most of the stuff that gets published in psychology is trivial. It will sink like a stone in a year. If there are errors or frauds in it, nobody cares enough to find out before it's forgotten anyway." (Ironically, Kamin was one of the first scientists to call attention to anomalies in Cyril Burt's published twin IQ data and to publicly accuse him of fraud [Kamin, 1974].)

Others who have speculated about the frequency of misconduct in scientific research have suggested that while outright fraud may indeed be rare, other forms of misconduct may be more common. Bailar (1986), for example, made a distinction between lying and deception: "In science, lying is condemned....Deliberate or careless deception short of lying, however, seems to be universally accepted and sometimes even promoted as part of the culture of science" (p. 259). Bailar gave as examples of accepted deception such practices as failure to explain to readers weaknesses in data, statistical testing of post hoc hypotheses, and fragmentary or selective reporting of findings.

Engler et al. (1987) also distinguished between the frequency of outright fraud and smaller deceptions. They assumed that premeditated, conscious research fraud is "probably rare" (p. 1383), but that competition and the constant pressure to produce results make it tempting "to forget an 'anomalous' result, to fail to scrutinize data in a collaborative project, to leave the collection of data to a fast-working colleague known to be 'sloppy,' or to accept gift coauthorship" (p. 1384).

Zuckerman (1977) speculated that the incidence of deviant behavior in science is an "iceberg phenomenon" (p. 99), with a lot more happening "under the water" than is detected. Zuckerman, a sociologist, said that for most forms of deviant behavior there are four potential sources of information about its incidence, none of which, unfortunately, are typically available concerning scientific misconduct: official statistics, self-reports, victimization statistics, and informants. Zuckerman believed that fraud in the sense of outright data fabrication is indeed rare, but that such ethical violations as theft of ideas or plagiarism, and secrecy about one's work, are more common.

Weinstein (1979) took the rather extreme and pessimistic view that policing in science is ineffective or nonexistent, that the norms of organized skepticism and universalism are not enforced and the norm of communality not practiced, and that replication seldom happens. She concluded, "There is no reason to believe [fraud is] less frequent than instances of false advertising and political coverup" (pp. 650-651).

Finally, Goodstein (1995) recently argued that whatever the real level of research misconduct may be, the perceived level is greater among laypeople (especially journalists and members of Congress) than among scientists themselves, because laypeople hold scientists to the unrealistic standards of the Myth of the Noble Scientist. What is harmless and acceptable fudging to the practicing scientist may be fraud to the outside observer.

Answers Based on Empirical Studies

Although the empirical evidence concerning the level of misconduct in scientific research is scanty and mostly flawed, studies do exist. Most of these studies are surveys with either low return rates or other serious sampling difficulties. One exception known to the present author is a Food and Drug Administration (FDA) report of 1,758 "essentially random" audits of clinical trials of experimental drugs, which found significant scientific misconduct in an astonishingly high 11% (Beardsley, 1988).

In a survey dealing with competitive behavior among scientists, Hagstrom (1974) surveyed 1,309 academic mathematicians, statisticians, physicists, chemists, and biologists. A quarter of the respondents reported that "another scientist [had] published results [I] published earlier without referring to [my] work" (p. 9) (and "probably knew" of the earlier work). The return rate on Hagstrom's survey is impossible to tell from the information in the article. At one point he reported an "89% response" (p. 2) to mailed questionnaires and follow-up telephone interviews. Later, however, he said that some of the data were based only on the 72% of scientists in the sample who completed the questionnaire (and the question about others not referring to one's earlier work was not asked in the interview). So the response rate could have been as small as 72% of 89%, or 64%. Unfortunately for the purposes of this review, Hagstrom's comparatively sound survey did not deal with misconduct issues other than failure to cite others' work.

Mahoney and Kimper (1976) did a survey of 400 scientists in physics, biology, psychology, and sociology. Of these, 23% (physics) to 57% (sociology) knew of someone in their field suppressing or discarding "negative" data, and 31% (physics) to 57% (biology) knew of someone fabricating data. Of course, these findings, especially those concerning fabrication, could have been accounted for by a very few actual cases, since these incidents tend to be widely publicized when discovered.

Mahoney and Kimper (1976) also asked respondents to estimate the percent of scientists in their fields who had suppressed data (2% in biology to 26% in psychology), faked data (0.4% in biology to 4% in sociology), plagiarized others' ideas (3% in biology to 23% in psychology), plagiarized others' writings (1% in biology to 8% in sociology), and misrepresented methodology (2% in biology to 22% in sociology). All estimates were considerably higher in psychology/sociology than in physics/biology, a striking phenomenon that Mahoney and Kimper chose not to discuss further, leaving open the question of the degree to which the higher percentages of misconduct reported by sociologists and psychologists really reflect a higher incidence of unethical research conduct in those fields, as opposed to a greater sensitivity of scientists in those fields to such misconduct. In any case, the value of Mahoney and Kimper's findings is severely compromised by their 19% return rate, a fact that they acknowledged and that may have explained their reluctance to discuss their meaning in any detail.

St. James-Roberts (1976a, 1976b) published a survey form in the magazine New Scientist, and so obtained a self-selected sample of 199 usable responses, representing an unknown percentage of scientist readers. Personal knowledge of data massaging was reported by 74% of respondents, experiment rigging by 17%, complete fabrication of experimental data by 7%, and deliberate misinterpretation by 2%. However, it is obvious that, as Zuckerman (1977) commented, the "[samples were] so small and unreliable as to preclude serious analysis" (p. 100). St. James-Roberts' results can hardly be considered more than a curiosity; the fact that they have been cited fairly frequently underscores the paucity of empirical knowledge in this area.

Petersdorf (1986; 1989) reported two different studies of medical students. His 1986 report cited a study (Barrett, 1985) of 400 medical students by doctors at two Chicago medical schools in which 88% of the respondents admitted to cheating as premeds, and that the majority said they continued to cheat in medical school. While this was not directly a study of researchers, Petersdorf assumed the phenomenon infects medical research, as can be seen from the title of his article: "The Pathogenesis of Fraud in Medical Science." Petersdorf concluded that medical school culture promotes cheating in medical research, and that medical science is "too competitive, too big, too entrepreneurial, and too much bent on winning" (p. 253).

Petersdorf (1989) cited a survey of medical students at an unnamed school in which 33-48% (depending on class) admitted cheating as undergrads, while 17% admitted cheating as medical students, including plagiarizing and faking lab results and results of research projects. This survey had a 71% response rate, but no other methodological details or references are provided. Petersdorf, the president of the American Association of Medical Colleges, concluded that "[it] is clear that fraud is a major affliction of science and medicine" (p. 121).

One must be careful not to reach conclusions about the relative rate of misconduct in various fields based on the number of cases which become public. Keeping this caveat in mind, Petersdorf's concerns about the level of misconduct in biomedicine are compatible with Woolf's (1988) analysis of characteristics of 26 cases of scientific misconduct that surfaced between 1980 and 1987. Based on "public sources such as news reports and scientific papers; Freedom of Information Act documents; telephone and personal interviews; and a few limited surveys" (p. 37), Woolf found that 2 cases were in chemistry/biochemistry, 1 in physiology, 2 in psychology, and the remaining 21 in the biomedical sciences. Of these 21, 17 involved M.D.s, not Ph.D.s. Overall, 22 of 26 either had M.D.s and/or were connected to medical schools, hospitals, or medical research institutions. Woolf also noted that for the most part prestigious schools/hospitals were involved, and perpetrators were often from similarly well known schools.

Fox (1990) also noted that most known cases of misconduct have been in natural and biological sciences, and speculated that "compared to the natural sciences, research in the social sciences is lower-stake, less competitive, and more poorly funded--less 'hot.' Accordingly, in social science, malpractice might be regarded as less consequential--less likely to be reported, if detected, and less likely to be indicted, if reported" (p. 68). Combining Fox's speculations with Mahoney and Kimper's (1976) findings reported above, of much higher estimates of misconduct in their fields by social scientists than by natural scientists, it is possible that a higher apparent rate of misconduct in the natural/medical sciences may mask a higher actual rate in the social sciences.

Tangney (1987) distributed questionnaires to scientists in the physical, biological, and behavioral sciences at one unnamed university. She reported that 32% of respondents suspected a colleague of falsifying data, and 32% suspected a colleague of plagiarism. Tangney reported she distributed 1,100 questionnaires, of which 245 were returned, for a mere 22% return rate; however, she believed the actual return rate was higher, because she provided surplus questionnaires to be distributed by department staff members. Tangney provided no other methodological details, which, combined with her low return rate, makes the results very difficult to interpret. However, making the most conservative possible estimate out of Tangney's numbers (i.e., the lowest possible rate of misconduct) requires two assumptions: that the 1,100 questionnaires ended up in the hands of 1,100 different scientists, and that all 855 nonrespondents would have reported no data falsification or plagiarism. Making these extremely conservative assumptions reduces Tangney's estimated rate of misconduct quite a bit, but still yields a 7% rate of suspicion of data falsification and plagiarism.

Kalichman and Friedman (1992) conducted a survey of 2,010 biomedical trainees at UC San Diego, obtaining a mere 27% response rate. Of these, 21% had been involved in "gift authorship," 10% had firsthand knowledge of data fabrication, and 36% reported personal knowledge of misconduct including inappropriate authorship or fraud. Further, 15% admitted to committing some type of personal misconduct (i.e., cheating on a test, modifying or making up research results, plagiarism), and 15% indicated willingness to consider misrepresenting data to publish a paper or secure a grant.

A recent article by Swazey et al. (1993) reported the results of a survey of 2,000 graduate students and 2,000 faculty in microbiology, civil engineering, sociology, and chemistry. Respondents were asked "Have you observed or had other direct evidence of any of the following types of misconduct? Please indicate the number of graduate students and faculty members whose misconduct you have observed/experienced" (p. 544). Faculty were asked to respond with respect to their current department, during the past 5 years. The results show that 6 to 9% of both students and faculty had direct knowledge of faculty who have plagiarized or falsified data; 16% of graduate students knew of other students falsifying data; 33% of faculty had seen graduate student plagiarism (more than 40% in civil engineering and sociology); 43% of faculty knew of colleagues misusing university resources; 29% of faculty knew of inappropriate authorship of research papers; 22% of faculty reported colleagues overlooking sloppy data use; and 15% knew of suppression of data that contradict researchers' own previous work. Swazey et al.'s (1993) survey reported response rates of 72% and 59%, respectively, for graduate students and faculty, the highest found by the present author among the studies attempting to use survey methods to estimate the rate of research misconduct in science.

Reducing the Amount of Misconduct in Scientific Research

Over the past decade or so, many authors have proposed reforms they believe would reduce the incidence of research misconduct in science. These proposed reforms range from obvious and comparatively minor changes to radical, even revolutionary actions. Milder suggestions for reform include making all coauthors jointly responsible for the accuracy of articles (e.g., Fox, 1994; Freedman, 1992), producing formal written guidelines for researchers (e.g., Stewart & Feder, 1987), or establishing formal organizational mechanisms to investigate alleged misconduct (e.g., Institute of Medicine, 1989; Kohn, 1986). More extreme reforms include such suggestions as shifting the burden of financial support for scientific research from government to private patrons (Broad & Wade, 1982).

The authors of proposed reforms have come from varied professional fields, and some of their ideas for reform clearly have been related to their backgrounds. Bell (1992), a professor of economics, focused on abuses related to funding systems in "big science," and proposed changes in these systems as solutions, as well as legislation to prevent financial conflicts of interest and to protect whistleblowers. Journalists Broad and Wade (1982) were perhaps the most radical; in addition to privatizing research funding, they also recommended reducing the number of journals by ending the practices of levying page charges and of government subsidies to libraries. (In fairness, it should be noted that they also made less radical proposals like requiring all listed authors to have made definably major contributions.)

Journal editors Fox (1994) and Freedman (1992) focused on the role of journals. Fox recommended such changes as alerting reviewers to look for suspicious results (e.g., too perfect, not enough variance, etc.), requiring authors to make raw data available, and holding all authors jointly responsible for work presented as collaborative, to prevent ex-post-facto disclaimers and diffusion and denial of responsibility. Freedman recommended that authors tell the complete story of their research, in the journalistic sense: "[T]he who, what, where, how, and why of an inquiry must be reported" (p. 194).

Other authors also focused on fragmentary reforms based on their professional roles. Statisticians Kromrey (1993) and Ellenberg (1983) recommended incorporating ethical considerations into graduate research and statistics coursework and establishing formal guidelines for statisticians, respectively. Mishkin (1988), an attorney, pointed out that "an important element in due process, often neglected in discussions of scientific misconduct, is adequate notice concerning acceptable standards of conduct and sanctions that may be imposed for failure to meet those standards" (p. 1932). In earlier times, Mishkin continued, these standards were passed on informally from senior scientists working closely with students. Today, however, these close relationships are often lacking; consequently, written research standards are necessary, and these must be incorporated into faculty contracts and student and faculty handbooks, discussed with incoming students and faculty, and incorporated into courses.

Hilgartner (1990) wrote from the perspective of an Institutional Review Board (IRB) member, and presented an excellent organizing schema for categorizing the orientations of various methods for reducing misconduct: (a) law enforcement, which focuses on detection, deterrence, and punishment; (b) oversight, which focuses on scrutiny of results, data, and practices; (c) education, which focuses on training and professional socialization of researchers; and (d) reward system, which focuses on changing the system to reduce incentives to cheat--that is, changing the "rules of the game" (p. 2) regarding academic appointments, promotions, grants, and so forth. Hilgartner limited himself to recommendations involving the IRB role--safeguarding human subjects, ensuring that researchers are thinking about recordkeeping and data-retention issues, and discouraging financial conflicts of interest.

In contrast to the fragmentary solutions reviewed above, there are a few published instances of sweeping, detailed recommendations for reforms to reduce the level of research misconduct. The two best examples located by the present author are from an Institute of Medicine Committee on the Responsible Conduct of Research (Institute of Medicine, 1989), and from Hersen and Miller (1992).

The committee that produced the Institute of Medicine (1989) report made several recommendations for action by the National Institutes of Health (NIH), by universities and other research centers, and by professional and scientific organizations and journals. For the NIH, the recommendations included (a) the establishment of an NIH office to promote responsible research practices; (b) a requirement that grantee and applicant institutions adopt policies and procedures to encourage responsible research practices, and a requirement that applicants affirm familiarity with these policies; (c) that the NIH not implement random data audits of investigator-initiated research;[3] and (d) the adoption of a policy to limit the number of publications that can be considered as part of a grant application.

For universities and other research centers, the committee's recommendations were that (a) they provide formal instruction in good research practices by incorporating such instruction into various places in the curriculum, as well as by formal courses; (b) they designate certain administrators and faculty members to promote responsible research practices, and that universities provide mediation and counseling services for faculty, staff, and students who wish to express concerns about questionable training or practices; (c) they modify incentives and academic guidelines to reduce pressure for

---

[3]The reasons given for this recommendation are that the nature of research in basic science is such that only subject-matter experts can properly interpret raw data, that the need for large numbers of such experts could make costs prohibitive, that an audit program could unduly diminish spontaneity and creativity in the research environment, and that it would not detect all instances of fraud (e.g., raw data fabrication).

excessive publication; (d) they monitor supervisory and training practices to ensure adequate oversight for young scientists; and (e) they adopt clear authorship policies to improve publication practices.

For professional and scientific organizations and journals, among the committee's recommendations were (a) organizations develop educational and training activities and materials to improve research integrity, and assist universities in identifying substandard practices; and (b) journals develop policies to promote responsible authorship policies, including procedures for responding to allegations or indications of misconduct.

Hersen and Miller (1992) endorsed wholly the recommendations made in Institute of Medicine (1989). Additionally, they recommended that (a) promotion committees should consider the quality of candidates' published work rather than its quantity; (b) journal editors' bias against negative results should be reversed, that is, they should encourage publication of studies that refute existing notions; (c) grant procurement abilities should not unduly influence promotion decisions; (d) chairpeople, mentors, and other leaders should be more nurturing of students and less concerned about their securing independent funding; and (e) researchers funded by commercial enterprises to evaluate products should submit their work to colleagues for independent review.

Table 2 contains a compendium of suggested reforms the author has encountered in the literature on this topic.

One Suggested Reform: Prevention of Misconduct Through Education

As perusal of the various suggestions for reform discussed above will demonstrate, one frequently suggested angle of attack on research misconduct is prevention through instruction of graduate students in research ethics. Specific means of instruction

Table 2

Compendium of Reforms Suggested to Reduce Research Misconduct

Type of Misconduct	Authors
1. Improved university oversight (including the mentoring process).	Bell, 1992; Hersen & Miller, 1992; Hilgartner, 1990; Institute of Medicine, 1989; Kohn, 1986; Mishkin, 1988; National Academy of Sciences, National Academy of Engineering, & Institute of Medicine, 1992; Petersdorf, 1986, 1989; Relman, 1989; Steneck, 1984, 1993; Woolf, 1981
2. Improved oversight by grant agencies.	Institute of Medicine, 1989; Relman, 1989
3. Formal guidelines for researchers and other professionals.	Ellenberg, 1983; Institute of Medicine, 1989; Kohn, 1986; Petersdorf, 1989; Steneck, 1993; Stewart & Feder, 1987
4. Formal education in research ethics.	Institute of Medicine, 1989; Kromrey, 1993; Mishkin, 1988; National Academy of Sciences et al., 1992; Steneck, 1984, 1993
5. Authorship reforms (e.g., making all authors responsible for entire article, elimination of "gift.")	Broad & Wade, 1982; Fox, 1994; Freedman, 1992; Institute of Medicine, 1989; Kohn, 1986
6. Journal, grant reviewers alert for signs of fraud.	Fox, 1994; Petersdorf, 1986
7. Require authors to keep raw data available.	Fox, 1994; Freedman, 1992; Woolf, 1981
8. Journals print retractions when fraud proven.	Fox, 1994; Woolf, 1981
9. Journals end bias against publishing "negative" results.	Hersen & Miller, 1992; Mahoney, 1976
10. Articles written in jounalistic style, telling "complete story" of research.	Freedman, 1992
11. Social structure of science encouraging quality of research over quantity.	Broad & Wade, 1982; Hersen & Miller, 1992; Institute of Medicine, 1989; Relman, 1989
12. Formal procedures for dealing with misconduct once it is alleged.	Institute of Medicine, 1989; Kohn,1986; Petersdorf, 1986, 1989; National Academy of Sciences et al., 1992

(table continues)

to make sure students practice ethical responsibility, and to encourage a healthy research atmosphere. In a later article, Steneck (1993) again called for formal and informal discussion of issues by faculty in day-to-day research settings, departmental and university forums, lecture series, and orientation programs, and formal instruction. Steneck also described efforts to promote discussion of ethics at Michigan, Harvard, Colorado, and Texas.

The Present Situation Regarding Ethics Education in Graduate Schools

It is not known how much ethics education is actually happening in graduate schools. In a recent article, Steneck (1994) noted that "by March 1989 most of the major and about half of the middle range research universities reportedly had adopted scientific misconduct policies. A much smaller percentage of universities (no estimates are available at the present time) have undertaken systematic efforts to foster research integrity" (p. 315), and that "...a few universities have taken some initiatives by beginning new courses..." (p. 325). Steneck believed that most of these efforts are driven by National Science Foundation and Public Health Service requirements, and that most universities, rather than instituting further efforts, are merely "meeting current requirements and waiting to see what happens" (p. 325). Furthermore, just as the degree to which universities are teaching research ethics to their students is unknown, so is the relative emphasis on different methods of ethics instruction. How many universities have instituted and are requiring separate courses in research ethics? How many are incorporating ethics into other graduate courses? How many are requiring students to read handbooks or policy statements concerning misconduct? How many rely on students' mentors to transmit this information informally? All of these are unknown. If we know little about how many universities are providing ethics education to graduate students, we know even less about how they are doing it.

The Efficacy of Ethics Education

How effective might education be in reducing research misconduct? Attempting to change the ethical standards of young adults concerning research misconduct would seem to involve two possible avenues of effectiveness, depending on whether it is seen as simply a matter of education or one of moral development. If it is merely education, then it involves simply teaching the customs and ethos of science to students who may not know or understand them. The second possibility is more problematic--is it possible, and to what degree, to teach adults a higher standard of ethical behavior? Research indicates the answer may be "yes, it is possible."

Rest (1988), in a review article, claimed that "deliberate educational attempts (formal curriculum) to influence awareness of moral problems and to influence the reasoning/judgment process can be demonstrated to be effective" (p. 23), and that "studies link moral perception and moral judgment with actual, real-life behavior" (p. 23). Rest concludes that the literature shows that

> ...students in professional schools are in a very important formative period of ethical development, that formal schooling is a powerful catalyst to ethical development...and that even rather modest and low cost educational interventions can produce significant results. (p. 23)

Leming (1981) reviewed studies on the effectiveness of moral/value education, with 27 studies that focused on a moral development approach. The studies were required to have a sound experimental design and to have involved group studies conducted in a classroom setting. All used the Kohlberg (1978) approach involving (a) exposure to the next higher stage of reasoning, (b) exposure to situations posing problems and contradictions for the current moral structure, and (c) an atmosphere of interchange in which conflicting moral views are compared. In all studies, the dependent variable was the stage of moral reasoning. Leming's review found two studies using college students, both of which had statistically significant results in favor of the moral education group. Overall, 22 of 27 studies found such differences, including 11 of 15 using students in grades 7-12.

Schlaefli, Rest, and Thoma (1985), in a meta-analysis, reviewed 55 studies of education interventions designed to stimulate development in moral judgment. Their principal findings: Dilemma discussion and psychological development programs produce modest overall effect sizes, treatments of 3-12 weeks are optimal, and programs with adults (age 24 and older) produce larger effect sizes than with younger subjects. Schlaefli et al. found nine samples using adults. In these nine, the mean pre-post effect size (weighted by sample size) was 0.61. Six samples measured a control group; the control groups' mean pre-post effect size was -0.13.

Taken together, it would appear the research evidence shows that it is at least possible for education to have a positive impact on the moral judgment of young adults.

Ethics Education Concerning Research Misconduct Specifically

The above studies deal in the general area of moral judgment, not specifically with research misconduct. How effective is ethics education concerning research misconduct for graduate students? Does it really affect students' standards, much less their behavior? The author knows of no published studies of evaluation of efforts to provide research ethics education. Perhaps the closest approximation is that of Self, Wolinsky, and Baldwin (1989), who studied the effects of incorporating a medical ethics (not research ethics) course into the medical school curriculum at Texas A&M University. The comparison group, oddly enough, was a group of veterinary students.

The study found a difference in favor of the students taught ethics, as measured by the score on Gibbs' Sociomoral Reflection Measure (a paper-and-pencil version of the Kohlberg Moral Judgment Interview), with a standard mean difference effect size estimated by the present author of around 1.0, and statistically significant at $p < .0001$. Interestingly, in Self et al.'s pre-post design, the treatment group average score on the Gibbs measure stayed the same over time, while the comparison group average declined.

Kalichman and Friedman (1992), in the survey of 2,010 biomedical trainees described earlier in the present review, noted that some prior training in research ethics, either through coursework or mentorship, was reported by 76% of survey returnees; however, no association was found between having had courses or discussions about ethical issues and reports of observing, participating in, or willingness to consider acts of misconduct. It is difficult to be confident in these conclusions, however, because of the 27% response rate to the survey.

Summary

To summarize, the author's review of the literature related to research ethics leads to the following conclusions:

1. There is not universal agreement as to what constitutes research misconduct, although some actions (e.g., fabricating data) are presumably considered wrong by all scientists, so strongly do they violate norms.

2. There is much speculation but little knowledge about what causes scientists to commit unethical acts.

3. Likewise, there is little solid empirical knowledge about the amount of misconduct in scientific research. Studies have reached varying conclusions, but a

recent, reasonably sound survey (Swazey et al., 1993) found that 6-9% of graduate students and faculty in four disciplines had direct knowledge of faculty plagiarising or falsifying data.

4. Several ways to reduce the level of misconduct have been suggested, including prevention of misconduct through educating graduate students in ethical conduct as part of their training.

5. Little is known about the extent or nature of research ethics education of students in Ph.D. programs, including the percentage of programs that have some type of formal research ethics education, how it is done, or how effective it is.

CHAPTER III

PROCEDURES

The present study had two parts: (a) a survey of department heads in three selected disciplines to determine the present state of ethics education in those disciplines, and (b) a survey of graduate students in the same departments as in the department head survey, to assess attitudes toward various forms of possible misconduct, ranging from the relatively innocuous (e.g., reporting statistical significance while knowingly failing to report a small effect size) to the most serious (e.g., making up data). Respondents were asked to rate the seriousness of each type of misconduct listed; individual ratings were then averaged to get a mean seriousness rating, which was considered a measure of the overall rigor of the student's ethical standards.

The disciplines were chosen to represent three of the four types in the first two dimensions (hard-soft, pure-applied) of Biglan's (1973) typology of academic disciplines. The disciplines of physiology, psychology, and mechanical engineering were chosen, because they meet the selection criteria, and because they occupy the same quadrants as the disciplines surveyed by Swazey et al. (1993), thus building on their work. (The "soft-applied" quadrant was excluded, because a relatively low proportion of people with Ph. D.s in those fields--for example, education--are involved in research careers.)

Department Head Survey

All heads of departments offering the Ph.D. in their respective disciplines were surveyed. The departments were those that participated in a national survey of graduate education conducted in 1993 by the National Research Council (Goldberger, Maher, & Flattau, 1995). That survey included any university that had produced at least three Ph.D.s in 1988-90 and at least one in 1991, in addition to any program that had achieved a certain minimum rating in a similar study conducted in 1982. In all, 300 universities were invited to participate, 284 agreed to do so, and 274 actually did (91%).[4] The numbers of departments thus listed by Goldberger et al. (1995) are 140 in biology, 185 in psychology, and 110 in mechanical engineering. Twelve departments in psychology were excluded from the present study, because they award degrees only in clinical, counseling, industrial, and/or organizational psychology; many, if not most, people who receive degrees in these areas are not employed in research.[5] (Goldberger et al. [1995] was also the source for information on the size and program effectiveness ratings of the departments surveyed.)

The survey titled "Department Head Questionnaire" (see Appendix A) was mailed to the heads of the selected departments, using names and addresses from published sources (American Psychological Association, 1995; Peterson's Guides, Inc., 1995a, 1995b). Accompanying the survey form was a cover letter, on USU Psychology Department letterhead, explaining the study (Appendix A). In addition to completing the questionnaire, department heads were asked to provide the names of the graduate

---

[4]Eligible universities that declined to participate were California Institute of Integral Studies, Graduate Theological Union, The Juilliard School, Indiana State University-Terre Haute, Long Island University-Brooklyn, Manhattan College, Marquette University, Middlebury College, Midwestern Baptist Theological Seminary, New School for Social Research, Nova University, Peabody Institute-Johns Hopkins, Southwestern Baptist Theological Seminary, U. S. International University, Villanova University, and Wright Institute. Universities that agreed to participate, but did not, were Caribbean Center for Advanced Studies, Cornell University Medical School, Cleveland State University, University of Dallas, Depaul University, Louisiana Technical, Memphis State University, Oregon Health Sciences, South Dakota State University, and Wright State University.

[5]Excluded universities in psychology are Arizona State University, Bryn Mawr University, California School of Professional Psychology (CSPP) at Alameda, CSPP at Fresno, CSPP at Los Angeles, CSPP at San Diego, Duquesne University, Fielding Institute, University of Nebraska-Lincoln, Pacific Graduate School, University of Tulsa, University of Health Sciences, and University of Texas-Southwestern Medical School.

students in their department.[6] Questionnaires were sent to a total of 390 department heads--107 in mechanical engineering, 130 in physiology, and 153 in psychology. Nonrespondents to the original mailing were sent a followup letter about 2 weeks later; those who still did not respond were sent a second followup, along with another copy of the questionnaire, about 3 weeks after the first followup. (See Appendix A for copies of the followup letters.) All envelopes were hand-addressed and hand-stamped.

Graduate Student Survey

When this study was planned, it was without any knowledge of what departments in the various discipline were actually doing with respect to research ethics education; indeed, finding that out was exactly the aim of the first phase of the study, the department head questionnaire. The tentative plan for the second phase, as reflected in research questions 3 and 4, was to compare the ratings of students from departments offering formal research ethics educations to ratings of students from departments not offering such education, and within the former type, to compare the ratings of students where such education was required to those of students where it was not required. However, the plan was flexible; if it turned out that the pattern of responses from the department heads was such that a different grouping was preferable, then adjustments would be made. As it turned out, only four (5%) psychology departments and only 10 (15%) physiology departments could be classified

---

[6]Psychology department heads were asked to exclude clinical psychology students from the list, because such a high percentage of them do not plan research careers. It was decided not to ask them to exclude students in organizational, industrial, or counseling, because (a) compared to clinical students, they are relatively few in number, and (b) it was feared that the department heads might decide that distinguishing between students in so many different areas was too much trouble, and toss the whole survey in the trash. Students in all areas were asked if they planned research careers--only 4% said "no," and their survey responses were no different than those from other respondents.

as not offering any formal education in research ethics (defined as checking <u>only</u> "informal learning through students' ongoing contact with faculty members" or "none"). It was believed that the way to get the maximum possible information from phase 2--to best answer the conceptual question underlying research questions 3 and 4, which is "Is there evidence that research ethics education is effective?"--was to recast the research questions in such a way as to maximize the power of the study. Accordingly, rather than sampling students from departments "with" and "without" formal ethics education, it was decided to sample students from departments that place a relatively high degree of emphasis on research ethics education versus departments that place a relatively low degree of emphasis on research ethics education, based on department heads' responses to their questionnaire. The new research question, replacing the original questions 3 and 4, is thus: "What is the difference between students in departments that place a high degree of emphasis on research ethics education, and those in departments that place a low degree of emphasis on research ethics education, with respect to how they judge the seriousness of various kinds of possible misconduct?"

The following procedure was used to divide departments into high- and low-emphasis groups. The nine possible means of providing research ethics education listed on the department head questionnaire (all possible responses on question 1 except "other") were rank ordered by degree of emphasis indicated by each (in the judgment of the author and of the committee chair). These rankings were, from highest emphasis to lowest:

1. A <u>required</u> class specifically dedicated to teaching standards of research ethics.
2. Part of <u>required</u> class(es) primarily dedicated to other content areas.

3. Attendance at required seminars, brown bags, or discussion groups on standards of research ethics.

4. Written handbooks or other written policy statements students are expected to read.

5. An elective class specifically dedicated to teaching standards of research ethics.

6. Part of elective class(es) primarily dedicated to other content areas.

7. Attendance at elective seminars, brown bags, or discussion groups on standards of research ethics.

8. Informal learning through students' ongoing contact with faculty members.

8. (tie) None--we do not explicitly teach reasearch ethics, as such.

With this rank ordering of degrees of emphasis in mind, grouping criteria were developed so as to give a reasonable number of departments (defined as 10 or more) in each high and low emphasis group[7] from which to sample students. For example, in defining "high emphasis," the number of departments within a discipline whose department heads marked the survey response defined as the highest emphasis, "A required class specifically dedicated to teaching standards of research ethics," was counted. If that yielded more than 10 departments, that became the criterion defining the high-emphasis group. If fewer than 10 departments met this criterion, the number of departments indicating the selection defined as the second-highest emphasis, "Part of required class(es) primarily dedicated to other content areas," was counted, and so on, until at least 10 departments qualified. The criteria ultimately selected for use were as follows:

---

[7]Because of rather large differences between disciplines (to be discussed in detail in the Results section), it was necessary to define low emphasis and high emphasis separately for each discipline. For example, the same definition of "high emphasis" used for physiology, which resulted in 37 departments being so classified, would have qualified only one mechanical engineering department.

Mechanical Engineering

Departments offering a required class specifically dedicated to research ethics, a required class partially dedicated to research ethics, seminars, brown bags, and so forth, at which attendance was required, or written handbooks or other written policy statements were considered high emphasis. Departments offering none of the above and offering informal learning through everyday faculty contact, or an elective seminar, or who indicated "none," were considered low emphasis.

Physiology

Departments offering a required class specifically dedicated to research ethics, or a required class partially dedicated to research ethics, were considered high emphasis. Departments offering neither of the above, nor a required seminar, brown bag, etc., nor a handbook or other written policy statement, but offering informal learning only, or an elective seminar, brown bag, and so forth, or who indicated "none," were considered low emphasis.

Psychology

Departments offering a required class specifically dedicated to research ethics, or a required class partially dedicated to research ethics, were considered high emphasis. Departments offering neither of the above, but offering an elective class entirely or partially dedicated to research ethics, or elective seminars, brown bags, and so forth, or informal learning only, or who indicated "none," were considered low emphasis.

This procedure resulted in 11 departments (6 of which supplied graduate student names) in the high emphasis group in mechanical engineering, and 25 (10 of which supplied names) in the low-emphasis group; 37 high-emphasis departments and 12 low-emphasis departments in physiology (with 21 and 8, respectively, supplying student names); and 51 high-emphasis departments and 10 low-emphasis departments in psychology (with 31 and 7, respectively, supplying student names). From each of the six groups (three disciplines, high vs. low emphasis in each), 100 students were randomly selected to receive surveys.

A copy of the graduate student questionnaire and an accompanying cover letter (both shown in Appendix B) were then sent to the random sample of graduate students in each discipline. Students were given a stamped envelope in which to return the questionnaire. The return envelope bore a code number, so follow-up requests could be sent (three weeks after the first mailing) to only those students who did not return the original questionnaire. A second follow-up, with another copy of the questionnaire enclosed, was mailed two weeks after the first follow-up. As with the survey of department heads, all envelopes were hand-addressed and hand-stamped.

Responses to each item on each questionnaire were coded into SPSS Windows data files (Norusis, 1993; SPSS, Inc., 1993). In addition, for the graduate student questionnaire, on which respondents rated on a 7-point Likert-type scale the seriousness of 44 different possible acts of misconduct, a mean seriousness rating across all 44 items was computed for each respondent. This value is considered a measure of the overall strictness of the student's ethical standards.

CHAPTER IV

RESULTS

Return Rates

Of the 390 department heads, 188 returned the survey, for a return rate of 48%. The return rates were 44 of 107 (41%) for mechanical engineering, 68 of 130 (52%) for physiology, and 77 of 153 (50%) for psychology. Of the 40 mechanical engineering departments returning the survey, 11 were eventually classified high emphasis, 25 low emphasis; the remaining 8 either had an emphasis level that fell between high and low, or were too late for student sampling, and were classified as neither. Of the 63 physiology departments, 37 were high emphasis, 12 low emphasis, and 19 neither. Of the 73 psychology departments, 51 were high emphasis, 10 low emphasis, and 16 neither.

Of the mechanical engineering departments returning the survey, 43% complied with the request to include a list of student names to be included in the sample for the graduate student survey; 60% of physiology departments and 62% of psychology departments also supplied student names. There was no relationship between department heads' willingness to supply names of students and their classification as high or low emphasis with respect to research ethics education. A summary of information concerning return rates, group designations, and numbers of student names is included in Table 3.

A mistake was made in sampling the 100 students from the low-emphasis physiology group. One department sent a list of student names on the same page as a list of faculty names, and the author entered the wrong list into the student name database. Consequently, 32 student names in the final sample were actually faculty,

Table 3

Summary of Information Concerning Return Rates, Grouping, and Student Name Samples From the Department Head Survey, by Discipline

Information	Mech. Engineering	Physiology	Psychology
Departments Surveyed	107	130	153
Returned Survey	44	68	77
Return Rate	41%	52%	50%
Supplied Student Names	17	38	45
Classified High Emphasis	6	21	31
Classified Low Emphasis	10	8	7
High Emphasis--Names	266	335	1435
Low Emphasis--Names	194	111	225

thus reducing the potential sample for the low-emphasis group by that number. In addition, another 18 surveys across the six groups were returned by the post office as "undeliverable"; presumably the departments had sent student lists that were not completely up to date. In all, then, the potential sample was reduced from 600 to 550, the number of students who actually received the survey.

Of the 550 students surveyed, 390 returned a completed survey, for a total return rate of 71%. Return rates by group are summarized in Table 4.

Table 4

Return Rate for the Graduate Student Survey, by Group

Group	Surveyed	Returns	Percent Returned
Mech. Eng.--High Emphasis	99	72	73
Mech. Eng.--Low Emphasis	97	53	55
Physiology--High Emphasis	91	70	77
Physiology--Low Emphasis	66	53	80
Psychology--High Emphasis	98	71	72
Psychology--Low Emphasis	99	71	72
Total	550	390	71

An Unexpected Problem and the Resulting Adjustment

While entering the data from the graduate student questionnaire into the computer file, an occasional pattern of responses that seemed clearly to be aberrant was noticed. Typically for this subgroup of respondents, every item would be marked "1" (i.e., the extreme end of the Likert scale, indicating "not at all an act of misconduct"). Occasionally, one would be marked with all 1s and 2s. These were "flagged" with an extra code on the computer file, so they could be kept track of and examined later.

Sixteen questionnaires were flagged as a result of this process. Upon examination of the computer file, which had not yet had respondents' names removed (followups were still being attempted), it was noticed that all but one of the 16 identified respondents had an Asian name, indicating that perhaps the instructions on the questionnaire had not been written in a sufficiently clear way that someone whose first language was not English could reliably understand them. One hypothesis that seems to fit this pattern of responses is that the respondent believed he or she was supposed to indicate how frequently occurring the behaviors listed in the items were, rather than how serious. This is something that should be addressed in follow-up studies, of course, but whatever the reason for it, it also presents a problem for the present study: What to do with data that seem clearly not to belong, especially when, as was the case, the aberrant responders were not evenly distributed across disciplines (almost all were in mechanical engineering or physiology) or across groups (high vs. low emphasis) within disciplines?

The following procedure was decided upon. First, the subjective element was removed, so as to prevent possible bias by the author. It was decided to flag any questionnaire on which the respondent rated "honest error" as equally or more serious misconduct than "fabrication of data," a response pattern that seemed to indicate that the respondent was not responding based on the seriousness dimension, but rather based on something else. Questionnaires were also flagged if they skipped either the "fabrication of data" item or the "honest error" item. An additional 17 questionnaires were flagged by this completely objective procedure (9 for rating honest error the same or worse than data fabrication, 8 for skipping one or both of the items), for a total of 33 aberrant respondents. These 33 were distributed as follows: 15 from mechanical engineering (8 from the high-emphasis group, 7 from the low-emphasis group); 15 from physiology (5 high emphasis, 10 low emphasis); and 3 from psychology (all low emphasis). Of the 33, 27 had names that appeared to the author to indicate that English may not have been their first language (23 Asian, 2 African, 1 Indian, 1 Spanish), lending support to the suspicion that perhaps something about the questionnaire, most likely the instructions, was insufficiently clear to nonnative English speakers. Among the remaining 6 flagged respondents, 1 answered "7" for all 44 items, 2 answered "1" for all items, and 3 skipped the "honest error" item (one of whom skipped the entire page 2 of the questionnaire). As a result of this procedure for removing aberrant responders, analyses of the graduate student questionnaire are based on 357 respondents, rather than the original 390.

Department Head Survey Results

The research questions to be answered by the department head questionnaire are shown below.

1. What are university departments, colleges, and central administrations doing to teach research ethics to graduate students? That is, what is the relative prevalence of such methods as formal classes, informal seminars or discussion groups, lectures, and written handbooks or other written policy statements? Where provided, is research ethics education handled at the department level, the college level, the university level, or some other? In how many departments that provide research ethics education is it required, rather than elective? How many departments do not provide research ethics education at all?

2. Are the relative frequencies of different methods of research ethics education different in different disciplines? Do they differ according to the quality rating or size of the department?

Tables 5 and 6 summarize the results of the department head questionnaire, showing percentages using each means of education, by discipline. Perhaps the most striking finding illustrated in Table 5 is the relative lack of emphasis on research ethics education by mechanical engineering departments. Only 11% of mechanical

Table 5

Percentage of Departments Using Various Methods of Research Ethics Education, by Discipline

Method	Mech. Engineering	Physiology	Psychology
Informal learning	96	85	94
Entire required class	2	46	16
Entire elective class	7	15	10
Part of required class	5	19	64
Part of elective class	7	7	27
Written handbooks or other statements	21	43	52
Required seminars or discussions	5	19	13
Elective seminars or discussions	16	15	22
None	16	6	0
Other	14	6	8
Informal learning only	46	15	5
Informal learning or "None" only	52	15	5
Some form required	11	62	74
Class required (part or entire)	7	60	71
N of departments	44	68	77

Table 6

Percentage of Departments Handling Research Ethics Education at Various Administrative Levels, by Discipline

Level	Mech. Engineering	Physiology	Psychology
Program	12	24	61
Department	71	53	79
College	24	26	5
Graduate school	39	46	25
University	22	11	12
Centralized only (i.e., checked "College," "Graduate school," or "University," and did not check "Program" or "Department"	7	29	0
N of departments	41	66	76

engineering departments require some formal research ethics education of their Ph.D. students; by contrast, 62% of physiology departments and 74% of psychology departments require formal education in research ethics. Similarly, only 7% of mechanical engineering departments require a class either entirely or partially devoted to research ethics, compared to 60% for physiology and 71% for psychology. Finally, 52% of mechanical engineering departments indicated either that informal learning through everyday contact with faculty was the only means of research ethics education available in their departments, or checked "none." By contrast, only 15% of physiology departments and only 5% of psychology departments indicated either informal learning only or none.

Interestingly, although more psychology departments than physiology departments require some form of research ethics education, and although more require it to be in the form of a class, far more physiology than psychology departments--46% to 16%--have a required class devoted entirely to research ethics. Psychology departments, by contrast, indicated far more frequently than physiology departments--64% to 19%--that they teach research ethics as part of a required class primarily devoted to other topics. This last finding may possibly mean that, as a discipline, physiology in fact devotes more total time to research ethics education than does psychology.

As for the question of which administrative levels are responsible for research ethics education of graduate students (Table 6), it appears that overall, the tendency is to do it in a fairly decentralized way, that is, at the program or department level. However, a sizable number of department heads, especially in physiology and mechanical engineering, report that research ethics education of their Ph.D. students is handled at a more centralized level than the program or department (i.e., at either the college, the graduate school, and/or the university level). This was especially true in physiology, where 29% of departments reported that the research ethics education of their students was handled exclusively at the college level or above. (Incidentally, none of the responses classified as "other" in Table 6 would have added to the "centralized only" category. Most such responses either said some variation of "individual faculty," or else said it was done at the department level, but a different department.)

Tables 7 and 8 summarize data bearing on the question of whether methods of research ethics education in use is related to either quality or size of the program. As can be seen in Tables 7 and 8, it appears that in all three disciplines, departments requiring some form of research ethics education have higher program effectiveness ratings,[8] and are larger (i.e., have more students) than departments that indicated they offer research ethics education either through informal faculty contact only or not at all. The small ns in some cells involved in the comparisons in Tables 7 and 8 make conclusions about those cells somewhat tenuous, however.

Finally, Tables C-1 to C-6 (Appendix C) show, for each discipline, percentages of departments using different means of research ethics education, broken down by quartile ranks on program effectiveness rating and by size of department. Tables C-1 to C-6 tell basically the same story as Tables 7 and 8--that larger and more effective programs tend to place a bit more emphasis on research ethics education--only in much more detailed form; they are provided in Appendix C for the sake of completeness.

Table 7

Mean Program Effectiveness Ratings, on a 0 (Low) to 5 (High) Scale, of Programs Having Different Degrees of Emphasis on Research Ethics Education, by Discipline

Type of Ethics Education	Mech. Eng.			Physiology			Psychology		
	Mean	SD	n	Mean	SD	n	Mean	SD	n
Informal only	2.7	0.7	20	3.1	0.4	10	2.7	0.3	4
Informal only or none	2.7	0.7	23	3.1	0.4	10	2.7	0.3	4
Some form required	3.3	0.6	5	3.3	0.6	40	2.9	0.6	57
Class required	3.0	0.1	3	3.3	0.6	39	2.9	0.6	55

[8]The source for the program effectiveness ratings was the same National Research Council survey (described in Goldberger et al., 1995) that was the source for the sampling frame of Ph.D. programs in mechanical engineering, physiology, and psychology.

Table 8

Mean Size (Number of Students) of Programs Having Different Degrees of Emphasis on Research Ethics Education, by Discipline

Type of Ethics Education	Mech. Eng. Mean	SD	n	Physiology Mean	SD	n	Psychology Mean	SD	n
Informal only	36.2	41.7	20	8.6	6.3	10	36.5	27.9	4
Informal only or None	33.3	39.5	23	8.6	6.3	10	36.5	27.9	4
Some form required	35.0	21.0	5	13.0	13.6	40	55.9	32.4	57
Class required	46.7	18.7	3	13.1	13.8	39	56.1	32.8	55

Graduate Student Survey Results

The research question to be answered by the graduate student survey was the recast question that replaced the original questions 3 and 4: What is the difference between students in departments that place a high degree of emphasis on research ethics education, and those in departments that place a low degree of emphasis on research ethics education, with respect to how they judge the seriousness of various kinds of possible misconduct?

To answer this question, the mean item rating across all 44 items was calculated for each student and added to the data file. (If one or more items was missing, the mean was calculated across all items actually answered.) In addition, four subscales were defined composed of items whose content related, respectively, to personal integrity, to methodology, to data analysis, or to publication. Items were assigned to subscales based on the judgment of the author and the committee chair, and no item was assigned to more than one subscale. Table 9 shows which items were assigned to each subscale.

Table 9

Graduate Student Survey Subscales and the Items Comprising Each

Item #	Description
Personal Integrity	
1.	Ad hominem attacks (i.e., criticizing a person instead of his/her work)
2.	Applying for funding to support work already done
6.	Failure to inform human subjects adequately
13.	False claims or commitments made in grant proposals
24.	Misrepresentation of publication status of an article (e.g., claiming it's "in press" when it has been submitted but not accepted)
25.	Mistreatment of human or animal subjects
32.	Retaliation against whistle-blowers
33.	Reviewing others' work unfairly (e.g., to sabotage a rival)
35.	False charges of plagiarism against others
37.	Using someone's research ideas without credit
38.	Performing research which ultimately has unintended bad consequences (e.g., work with genetically altered microorganisms unleashing a serious epidemic, or behavioral genetics research leading to discrimination)
42.	Using one's position to exploit or manipulate others
43.	Using university resources for outside consulting work
44.	Violation of privacy or confidentiality norms regarding subjects
Methodology	
3.	Carelessness in conducting experiments, including reading or recording data
5.	Fabrication of data
9.	Failure to make raw data available for re-analysis when requested
27.	Neglect or violation of methodological concerns and procedural precautions, (e.g., loosely following experimental protocol)
34.	Rigging experiments
41.	Using a poor research design
Data Analysis	
4.	Double checking only results that don't support one's hypothesis
14.	Selective deletion of "outlying" data points
16.	Honest error
17.	Incompetent data analysis
28.	Overlooking colleagues' use of flawed data, questionable interpretations, or other research transgressions
31.	Reporting the statistical significance of an effect while ignoring the magnitude of the effect
36.	Statistical testing of post hoc hypotheses (i.e., of hypotheses made after examining results)

(table continues)

Item #	Description
Publication	
7.	Failure to disclose weaknesses in data
8.	Failure to disclose weaknesses in research design
10.	Failure to present results that contradict one's previous research
11.	Failure to report results that do not support one's hypothesis
12.	Failure to publish until follow-up work is complete
15.	Giving only a cursory review to a paper submitted for publication, if it supports one's own theory
18.	Incomplete authorship (i.e., failure to credit someone who deserves coauthorship)
19.	Incomplete documentation of research procedures
20.	Intentional efforts to communicate false or misleading findings
21.	Intentional misinterpretation of results
22.	Misrepresentation of another's work in a citation
23.	Inaccurate representation of research or analysis procedures
26.	Breaking down the findings from a single piece of research into multiple papers
29.	Plagiarism (i.e., claiming another's work as one's own)
30.	Repeated publication of essentially the same content
39.	Underacknowledgment (e.g., failure to cite) of intellectual predecessors, rivals, colleagues
40.	Unjustifiable authorship (i.e., listing someone as an author who was not actually involved in doing the research or writing the article)

Cronbach's alpha coefficients for the four subscales were .84 for Personal Integrity (14 items), .70 for Methodology (6 items), .72 for Data Analysis (7 items), and .89 for Publication (17 items). Cronbach's alpha for the total 44-item scale was .94, indicating an extremely high degree of internal consistency. Correlations between subscales ranged from .58 to .77.

Tables 10-12 summarize the results for mechanical engineering, physiology, and psychology, respectively. For mechanical engineering, the total scale effect size[9] for students from the high-emphasis departments was 0.00 (i.e., there was no difference

[9]That is, the standard mean difference effect size--the mean score of the high emphasis students minus the mean score of the low emphasis students, divided by the standard deviation for the low emphasis students.

Table 10

Mean Scores of Students in High- versus Low-Emphasis Departments on the Graduate Student Questionnaire--Mechanical Engineering

	High emphasis (N = 65)		Low emphasis (N = 45)		
Scale	Mean	SD	Mean	SD	Effect size
Personal Integrity	5.5	0.9	5.5	1.2	0.03
Methodology	5.2	0.9	5.1	1.2	0.05
Data Analysis	3.8	1.0	3.9	1.3	-0.11
Publication	5.0	0.8	5.0	1.1	0.00
Total Scale	5.0	0.8	5.0	1.2	0.00

Table 11

Mean Scores of Students in High- versus Low-Emphasis Departments on the Graduate Student Questionnaire--Physiology

	High emphasis (N = 65)		Low emphasis (N = 43)		
Scale	Mean	SD	Mean	SD	Effect size
Personal Integrity	5.7	0.8	5.8	0.6	-0.19
Methodology	5.3	0.9	5.6	0.7	-0.46*
Data Analysis	3.9	1.0	4.1	0.8	-0.26
Publication	5.1	0.9	5.2	0.6	-0.23
Total Scale	5.1	0.8	5.3	0.5	-0.31

* $p < .05$

Table 12

Mean Scores of Students in High- versus Low-Emphasis Departments on the Graduate Student Questionnaire--Psychology

	High Emphasis (N = 71)		Low Emphasis (N = 68)		
Scale	Mean	SD	Mean	SD	Effect Size
Personal Integrity	5.8	0.6	5.6	0.6	0.23
Methodology	5.4	0.7	5.4	0.7	-0.10
Data Analysis	4.0	0.8	3.8	0.9	0.20
Publication	5.0	0.7	5.0	0.7	-0.04
Total Scale	5.1	0.6	5.1	0.6	0.09

at all between the high-emphasis mean and the low-emphasis mean). Subscale effect sizes ranged from -0.11 to 0.05. For physiology, the total scale effect size was -0.31--that is, the mean seriousness rating for the students from high-emphasis departments was 0.31 standard deviations *lower* than the mean for students from low-emphasis departments--with subscale effect sizes ranging from -0.46 to -0.19. For psychology, the total scale effect size was 0.09, with subscale effect sizes ranging from -0.10 to 0.23. None of the effect sizes (i.e., mean differences) was statistically significant at $p < .05$, as measured by a t test, with one exception: for physiology students, the methodology subscale effect size of -0.46 did reach statistical significance. Tables C-7 to C-9 (Appendix C) summarize the comparisons between students from high- and low-emphasis departments in mechanical engineering, physiology, and psychology, respectively, for each individual questionnaire item. Effect sizes are generally small, and only three are statistically significant at $p < .05$--

"Incompetent data analysis" and "Using a poor research design" were rated more serious by the students from low emphasis physiology departments, and "Using one's position to exploit or manipulate others" was rated more serious by students from high emphasis psychology departments (that is, two of three differences, both from physiology, were in the "wrong" direction). However, it should be kept in mind that with a total of 132 t tests being done (44 items times 3 disciplines), one would expect about seven alpha errors if the null hypothesis was true in each case (.05 x 132 = 6.6). Three statistically significant differences across 132 comparisons, therefore, may well mean nothing, and should not be relied upon without replication. In sum, it would appear that there is little difference in mean seriousness ratings between students from departments that place high emphasis on research ethics education, and students from departments that place low emphasis on research ethics education, in either mechanical engineering, physiology, or psychology.

Additional Question and Analysis

Although not specifically posed as a research question when the present study was originally proposed, prior to data collection an additional question was suggested by the committee chair and included by the author in analysis plans: What differences exist in students' mean seriousness ratings across disciplines? That is, do students in mechanical engineering, physiology, and psychology differ, by discipline, in the strictness of their beliefs about various possible acts of research misconduct? To answer this question, means and standard deviations were calculated by discipline. To test the statistical significance of the differences between means, a one-way analysis of variance was calculated; if the overall F value was statistically significant at $p < .05$, a post-hoc multiple comparisons test (Newman-Keuls) was planned, as recommended by Glass and Hopkins (1984, p. 376).

Table 13 shows the results of this analysis for the total scale and for the subscales. For the total scale, the mean score for physiology students was slightly higher than for psychology and mechanical engineering students, although not to a statistically significant degree. The value of the overall F test likewise did not reach statistical significance for any of the subscales.

For the sake of completeness, tests were also done on each individual item, even though neither the total scale nor the subscales showed any differences. Table C-10

Table 13

Mean Scores of Students on Subscales of the Graduate Student Questionnaire, by Discipline

	Mech. Engin. (N = 110)		Physiology (N = 108)		Psychology (N = 139)	
Scale	Mean	SD	Mean	SD	Mean	SD
Personal Integrity	5.5	1.0	5.7	0.8	5.7	0.6
Methodology	5.2	1.0	5.4	0.8	5.4	1.7
Data Analysis	3.9	1.1	4.0	0.9	3.9	0.9
Publication	5.0	1.0	5.1	0.8	5.0	0.7
Total Scale	5.0	1.0	5.2	0.7	5.1	0.6

Note. No differences among disciplines were statistically significant at $p < .05$.

(Appendix C) shows these individual item results. Only 12 of the 44 individual item means showed statistically significant differences across disciplines, with the relative strictness of students in the three disciplines showing no particular pattern; this fact, coupled with the fact that the items on which differences were found are scattered across the different broad areas of misconduct (i.e., those delineated in forming the subscales), makes the few differences found very difficult to interpret.

Results of Analyses Based on Respondents' Self-Reported Ethics Education Experience

All the results reported up to this point are based on comparisons of students from departments placing high emphasis on research ethics education to students from departments placing low emphasis on research ethics education, as reported by the department heads. The final question on the graduate student questionnaire asked the respondent to report his or her actual experience with research ethics education, by checking any or all that applied from a list of alternatives (e.g., a course specifically dedicated to teaching standards of research ethics, written handbooks, and so forth--see Appendix B). Responses to this item were used to assign graduate student respondents to one of two groups: those reporting they had had some formal education in research ethics, and those who reported having had no such education (i.e., either they checked "informal learning through ongoing contact with faculty members" only, or they checked "none"). Table 14 shows the results of this analysis.

The effect sizes for mechanical engineering and for psychology are quite small and not statistically significant. The effect sizes for physiology are reasonably large (0.41 for the total scale, 0.65 for the Data Analysis subscale), but still not statistically significant, although the small size (n = 14) of the “no formal ethics education” group results in low power in the statistical test.

Table 14

Mean Scores of Students Who Reported Having Some Versus No Formal Education in Research Ethics

	Some Formal Ethics Education			No Formal Ethics Education			
Discipline	Mean	SD	n	Mean	SD	n	Effect Size
Mechanical Engineering							
Personal Integrity	5.5	1.0		5.5	1.1		-.06
Methodology	5.1	1.0		5.2	1.1		-.08
Data Analysis	3.8	1.0		3.9	1.2		-.06
Publication	5.0	0.9		5.0	1.1		.00
Total Scale	5.0	0.9	57	5.0	1.0	53	-.05
Physiology							
Personal Integrity	5.7	0.7		5.5	1.0		.25
Methodology	5.4	0.8		5.3	0.7		.17
Data Analysis	4.1	1.0		3.6	0.7		.65
Publication	5.2	0.8		4.8	0.9		.42
Total Scale	5.2	0.7	94	4.9	0.7	14	.41
Psychology							
Personal Integrity	5.7	0.6		5.6	0.7		.15
Methodology	5.4	0.7		5.4	0.8		.02
Data Analysis	4.0	0.9		3.7	1.0		.31
Publication	5.0	0.7		4.9	0.8		.19
Total Scale	5.1	0.6	120	5.0	0.7	19	.19
All Disciplines Combined							
Personal Integrity	5.7	0.7		5.5	1.0		.12
Methodology	5.3	0.8		5.3	1.0		.08
Data Analysis	4.0	0.9		3.8	1.1		.15
Publication	5.1	0.8		4.9	1.0		.16
Total Scale	5.0	0.7	271	5.0	0.9	86	.14

Graduate student respondents were also asked to indicate the year in which they began study in their present program; 351 of the 357 complied. This information was then used to compute the variable "years in program," by subtracting it from 1996. If the ethics instruction that graduate students are receiving (in any form, including informal ongoing contact with mentors or other faculty) is effective, that effectiveness might be detectable as a positive correlation between years in school and mean seriousness ratings. This correlation was not found in the present sample; the correlation coefficient between the total 44-item scale and years in school was -.04. Correlations between years in school and the personal integrity, methodology, data analysis, and publication subscales ranged from -.07 to -.01. Correlations within each individual discipline were of similarly small magnitude.

Moreover, examination of mean scores broken down categorically by number of years in graduate school (i.e., comparison of mean seriousness ratings of students in graduate school for 1 year, 2 years, 3 years, etc.), revealed only miniscule differences, even between 1st- and 2nd-year students. In short, seriousness ratings were not related to number of years in graduate school.

A final attempt was made to explore the possibility of a low-threshold effect--that is, the possibility that ethics education <u>is</u> effective, but that the effects happen very early in the student's career. This was done by dividing students into four categories, based on whether or not they were 1st-year students and whether or not they reported having had any formal education in research ethics. If informal faculty contact is as effective as more formal methods at transmitting ethical standards, and if it happens relatively quickly, then the best place to find relatively lower standards might be among 1st-year students who have had no formal ethics education while enrolled in a graduate program. So students in that category might have lower mean scores than students in any of the other three, which would show up as an interaction effect in a 2 x 2 factorial ANOVA. This ANOVA analysis was done, but no statistically significant effect was found, although students in the "first year, no formal education" cell did in fact have the lowest mean of the four cells. Once again, the severely unbalanced cell sizes caused a loss in statistical power. Table 15 summarizes these results.

Nonresponse Bias Check

The target population for the department head survey was all large departments in physiology, psychology, and mechanical engineering in the United States. For the graduate student survey, the target population was all graduate students in departments placing either high or low emphasis on research ethics education. In both surveys, those who actually completed and returned the survey were, of course, a self-selected sample. A nonresponse bias check was therefore planned for both surveys, in order to check the possibility that nonrespondents were systematically different from respondents, thus affecting the results.

Table 15

Mean Seriousness Ratings by Years of Experience in Graduate School (First Year Versus Later Year), and by Whether or Not Students Had Formal Education in Research Ethics

	First Year Students			More Experienced Students		
Ethics Education	Mean	SD	n	Mean	SD	n
Some Formal Ethics Education	5.2	0.8	58	5.1	0.7	209
No Formal Ethics Education	4.8	1.1	18	5.0	0.9	66

As it turned out, for the graduate student survey, nonresponse bias was judged to be highly unlikely, for two reasons. First, the response rate was a very high 71%, near a rate which has been shown (Gough & Hall, 1977) to make nonrespondent bias exceedingly unlikely. Second, it should be noted that for bias to have an effect on the main question, not only would the relatively few nonrespondents have to be extremely different from the respondents in the strictness of their standards, but also, the nonrespondents from low-emphasis departments would have to be extremely different from nonrespondents from high-emphasis departments.

In spite of the extreme unlikelihood of nonresponse bias, an attempt was nevertheless made to contact over 20 (i.e., more than 10%) randomly selected nonrespondents by telephone, to persuade them to complete the survey. A person with a great deal of experience and training in doing phone surveys called the students' departments and asked to speak to the student, or to get a number where the student could be reached. If the student could not be reached immediately, repeated attempts were made, and messages were left asking the student to call back collect. In spite of these efforts, only three students were actually able to be contacted, and only one returned a survey. In light of the fact that nonrespondent bias is extremely unlikely to have affected the results in any case, no further attempts to contact nonrespondents were made.

Nonresponse bias was potentially of greater concern for the department head survey, because the return rate was only moderately high at 48%. Consequently, 16% of the nonrespondents were contacted and agreed to complete the survey by telephone--16% of the nonrespondents in mechanical engineering, 13% in physiology, and 18% in psychology. Results are shown in Table 16.

It was hypothesized that perhaps the results of the department head survey might be affected if departments that provide their students with little education in research ethics were less likely to respond. As Table 16 shows, if anything the opposite may be true--in every case where there is a substantial difference between respondents and nonrespondents, it is the nonrespondents whose departments appear to place more emphasis on research ethics education. What might the implications of this be, with respect to the research questions? Considering that the questions (and thus their answers) were purely descriptive in nature, any suggestion that nonrespondents may be different from respondents obviously calls the results in to question, in the sense that

Table 16

Results of the Nonrespondent Bias Check, by Discipline

	Mech. Engineering		Physiology		Psychology	
Method	Resp.	Nonresp.	Resp.	Nonresp.	Resp.	Nonresp.
Informal learning	96%	100%	85%	88%	94%	100%
Entire required class	2%	10%	46%	63%	16%	36%
Entire elective class	7%	0%	15%	13%	10%	7%
Part of required class	5%	10%	19%	63%	64%	86%
Part of elective class	7%	30%	7%	13%	27%	43%
Written handbooks or other statements	21%	70%	43%	38%	52%	71%
Required seminars or discussions	5%	20%	19%	38%	13%	50%
Elective seminars or discussions	16%	30%	15%	25%	22%	21%
None	16%	10%	6%	0%	0%	0%
Informal learning only	46%	10%	15%	0%	5%	7%
Informal only or None	52%	20%	15%	0%	5%	7%
Some form required	11%	30%	62%	100%	74%	93%
Class required	7%	20%	60%	100%	71%	93%
N of departments	44	10	68	8	77	14

the percentages reported in Table 5 and in the "respondents" column of Table 16 may not accurately reflect the true state of the world. The most straightforward interpretation of Table 16, with respect to the research questions, would be that it may be in fact that all three disciplines place more emphasis on research ethics education than one would conclude from looking at the data from respondents only.

However, in interpreting the results shown in Table 16, several things should be kept in mind: First, the number of nonrespondents who completed the survey by phone (i.e., those designated "nonrespondents" in the table) is relatively small. The differences between respondents and nonrespondents are thus suggestive, but not conclusive. Second, it may be that differences in the method of data collection (i.e., by phone rather than by a mail survey) may have affected the results. Finally and most importantly, it should be noted that the differences in favor of the nonrespondents exist across all three disciplines; therefore, a major conclusion of the department head survey--that mechanical engineering departments are providing far less ethics education to their Ph.D. students than are physiology and psychology departments--is unchanged. If anything, it may be strengthened.

CHAPTER V

SUMMARY, DISCUSSION, AND CONCLUSION

Summary

The present study had two parts: (a) a survey of department heads in three selected disciplines--mechanical engineering, physiology, and psychology--to determine the present state of ethics education of Ph.D. students in those disciplines, and (b) a survey of graduate students in a sample drawn from the same departments as in the department head survey, to assess attitudes toward various forms of possible misconduct, ranging from the relatively innocuous (e.g., reporting statistical significance while knowingly failing to report a small effect size) to the most serious (e.g., fabricating data). Respondents were asked to rate the seriousness of each type of misconduct listed on a 1 ("not at all an act of misconduct") to 7 ("an extremely serious act of misconduct") Likert-type scale; each respondent's individual item ratings were then averaged to get a mean seriousness rating, which was considered a measure of the overall strictness of the student's ethical standards.

To summarize the major results reported in the previous section, the following items were analyzed.

Return Rates

The study achieved fairly high return rates, 48% for the department head survey and 71% for the graduate student survey. The return rate of both surveys was noticeably lower for mechanical engineering than for physiology and psychology. A nonresponse bias check of department heads showed some sign of possible underestimation of emphasis on research ethics education in each discipline, but did nothing to change the conclusion that mechanical engineering departments place less emphasis on research ethics education than physiology or psychology departments. The response rate for graduate students was sufficiently high at 71% that it is believed that results were unbiased by nonresponse.

Department Head Survey Results

There are large differences between disciplines in the overall emphasis given to education in research ethics, with much more emphasis given in physiology and psychology than in mechanical engineering. A slightly higher percentage of psychology departments than physiology departments requires some form of formal education in research ethics (74% to 62%), but more physiology departments devote an entire required class exclusively to research ethics (46% to 16%), while more psychology departments teach research ethics as part of a required class primarily devoted to other topics (64% to 19%). By contrast, only 11% of mechanical engineering departments report requiring their doctoral students to have any form of research ethics education.

Graduate Student Survey Results--
Main Analyses

Evidence of any relationship between departments' emphasis on research ethics education and the strictness of students' ethical standards, as measured by mean seriousness ratings on the graduate student questionnaire, is very, very slight. Effect sizes for the total scale were very small--0.00, -0.31, and 0.09 for mechanical engineering, physiology, and psychology, respectively--and not statistically significant. Differences on individual subscales were likewise small and not statistically significant, with the exception of one subscale effect size for one discipline (physiology students on the methodology subscale), and that one was in the "wrong" direction (that is, students from low-emphasis departments were stricter).

Graduate Student Survey Results--Further Analyses

Mean ratings of the seriousness of acts of misconduct by students who either had or had not had some form of formal research ethics education, based on self-report, were compared. Effect sizes again were generally fairly small (-0.05, 0.41, and 0.19 in mechanical engineering, physiology, and psychology, respectively). In summarizing across students regardless of discipline, the effect size was 0.14. None of the effect sizes were statistically significant, but small ns in some cells resulted in low statistical power.

Mean seriousness ratings were not related to number of years respondents had been students in their graduate programs. This was true for the overall scale and for each subscale, and was also true for each discipline.

Discussion and Conclusion

The primary aims of this study, broadly stated, were (a) to find out what Ph.D. programs in mechanical engineering, physiology, and psychology were doing to educate their students concerning research ethics; and (b) to collect some evidence pertaining to the efficacy, or lack thereof, of formal education of Ph.D. students in the standards of research ethics. With respect to the first aim, this study was successful, and a good deal was learned. In psychology and physiology, very few departments rely solely on informal everyday contact with mentors and other faculty to inculcate standards of ethical behavior in the conduct of research. In mechanical engineering, by contrast, around half of the departments do rely solely on such informal methods, and very few require any formal instruction of any kind, even a seminar or discussion session. It cannot be known from this study, of course, whether this difference in emphasis on ethics education is characteristic of other disciplines in the same quadrants of Biglan's

(1973) typology--that is, if other disciplines in the "hard-applied" quadrant share mechanical engineering's relative lack of emphasis on ethics education, or if the higher emphasis in physiology and psychology is reflective of the "hard-pure" and "soft-pure" quadrants, respectively. Future studies should explore this issue.

With respect to the second broad aim of the present study--searching for evidence of the efficacy of research ethics education--one can only say that no such evidence was found. This in no way means, of course, that ethics education has been shown not to be effective. Failing to reject a null hypothesis never means the null was true, only that the study failed to show it was false. It is entirely possible, for example, that formal ethics education is very effective, but that (a) informal faculty communication is just as effective, and (b) all the effect takes place in the first few months or less of graduate school. Such a pattern of effectiveness would be beyond the capacity of the present (nonexperimental) study to detect, although further analysis of the results did show a pattern consistent with such an effect--1st-year students who reported having had no formal research ethics education had less strict standards than 1st-year students who did report having had such education, and also less strict standards than more experienced students. Future studies might explore this issue further.

Limitations

The present study has several limitations that affect the conclusiveness of the results. These limitations are discussed in this section.

One problem that needs to be addressed in follow-up studies is the questionnaire itself. Although experienced in designing questionnaires, and well aware of the need to keep the required reading proficiency level as low as possible, the present author naively thought that would not be an issue for a target population of Ph.D. students. The fact that many such students are not native English speakers, especially in the hard sciences, was overlooked in the planning of the study, and no international students were included in the dozen or so pilot respondents. This unfortunate oversight, it is believed, ultimately resulted in a clearly aberrant pattern of responses from a sizable number of respondents--around 7%--which in turn caused some loss of statistical power from reduced sample size. The instrument clearly needs to be revised before it is used in future studies.

Beyond this fairly straightforward revision, there are several other problems related to the graduate student survey instrument. Although the study was designed to investigate whether research ethics education was related to expressed standards of ethical strictness, rather than to actual behavior, it is nevertheless reasonable to say that while attitudes are valuable to measure in their own right, it is behavior--in the present case, ethical conduct--that ultimately matters. Thus, the question arises: Does the present study have anything to say about ethical conduct?

It is well known--indeed, in social psychology, it is almost a truism--that general attitude measures are not necessarily very good predictors of specific behavior, as exemplified by the classic study by LaPiere (1934), and summarized in an extensive review of the attitude-behavior research by Ajzen and Fishbein (1977). However, more recent studies suggest that measures of attitude toward specific behaviors may in fact be good predictors of those behaviors (Canary & Siebold, 1984). The items describing acts of possible misconduct contained on the graduate student survey were intended to be fairly specific in nature, and thus might be expected to be related to actual behavior. However, it is possible that a survey that attempted to assess strictness of ethical standards in a different way--perhaps through presenting more specific scenarios and asking students what they would do, rather than simply asking them to rate the degree to which actions represent serious misconduct--might be better related to actual behavior and thus truer measures of standards.

Two other potential limitations of the study are consequences of its design, which is, of necessity, correlational (in the sense that the independent variable, research ethics education, was not assigned by a central agent, a defining characteristic of an experimental design [Mohr, 1992]). As termed by Campbell and Stanley (1966) in their classic work, the present study was a (pre-experimental) Static Group Comparison. This design has, potentially, two types of internal validity problems: selection and mortality. Selection problems could arise if students were self-selecting into graduate programs based on the departments' ethics education policies (e.g., if students who have more stringent standards of research ethics before entering graduate school were selecting themselves into programs that have formal ethics education in place). Mortality problems could affect internal validity if, for example, students in departments with ethics education tended to drop out, and the dropouts were systematically different from the nondropouts in their beliefs about various types of misconduct. Neither of these possibilities is considered plausible by the present author--it is believed that students are unlikely to choose graduate programs based on whether or not formal education concerning research misconduct is provided, nor does ethics education seem likely to cause students to drop out of school. Consequently, neither is considered a serious source of threat to internal validity; thus, the correlational design itself is not a serious limitation.

Actually, the present design is somewhat stronger than a simple Static Group Comparison, because the treatment variable, ethics education, occurred independently in each individual program. As explained by Campbell and Stanley (1966), a correlational or ex post facto design is strengthened in cases where there are "numerous independent natural instances of [a treatment] and numerous ones of [no treatment]" (p. 64). If differences in the groups are measured, "the credibility of the hypothesis is strengthened in that it has survived a chance of disconfirmation" (p. 64). Further, Campbell and Stanley go on to say that such a design may approach the strength of a true experimental design in settings

> in which it seems plausible that exposure to [the treatment] was lawless, arbitrary, uncorrelated with prior conditions. Ideally these arbitrary exposure decisions will also be numerous and mutually independent....The causal interpretation of a...correlation depends upon both the presence of a compatible plausible causal hypothesis and the absence of plausible rival hypotheses to explain the correlation upon other grounds. (p. 65)

It is argued that the design of the present study meets Campbell and Stanley's conditions and thus does approach the strength of a true experiment, in terms of internal validity.

The second possible limitation that is a somewhat indirect consequence of the correlational design stems from the relative weakness of the "treatment" independent variable--that is, in the assignment of departments in each discipline to the high- and low-emphasis categories. As explained in the "Procedures" section, it was originally hoped, and tentatively planned, that the question of the effectiveness of research ethics education could be studied by comparing strictness standards of students from departments offering--or even requiring--research ethics education, to those of students from departments not offering such education. However, as was learned from the department head survey, very, very few departments in two of the disciplines, physiology and psychology, fail to offer formal research ethics education, and most require some formal coursework. In contrast, in mechanical engineering few departments require any form of research ethics education, and those that do tend to do so in what may be a relatively weak form (in a handbook or informal discussion group, for example, as opposed to a formal course).

Consequently, in all three disciplines, the difference between the high- and low-emphasis departments may not have been as great as one would hope, if one wished to maximize the chance of finding a detectable effect of research ethics education on students' beliefs concerning acts of possible misconduct. This is one of the sometimes unavoidable limitations inherent in doing nonexperimental research in real-world settings, and points up the need for universities to conduct solid evaluations of their efforts at research ethics education, including experimental studies. One possible design of such a study would be to randomly assign incoming graduate students to receive research ethics education in their first versus second semester, with standards assessed in some way at the end of the first semester--that is, at a point when one group has received the educational treatment while the other group has not.

Implications

What do the findings of the present study mean? What are the implications for practice and/or policy? What do they say about the adequacy of graduate programs' efforts at providing research ethics education? The major findings of the study are that there was little or no detectable effect of departments' degree of emphasis on research ethics education on the strictness of students' expressed beliefs concerning various acts that could possibly be construed as ethical misconduct, in any of three disciplines studied; moreover, the strictness of mechanical engineering students' standards is indistinguishable from that of students in physiology and psychology, even though, as a group, the department heads in their discipline report far less emphasis on formal research ethics education than do the department heads in physiology and psychology.

Given these findings, the question naturally arises: Considering that bodies of knowledge in most if not all disciplines seem to be growing at ever-faster rates, and considering that graduate programs are coming under increasing pressure to move students quickly through to degree completion, are administrators justified in interpreting the present findings as evidence that they would be justified in reducing or even eliminating the time and effort they spend on formal education in research ethics? The present author believes the answer is "no," for several reasons. First, as mentioned above, the logic of null hypothesis testing of samples is such that a "no effect" or "no difference" finding can never be interpreted as confirmation of the null, but only as a failure to reject it in one particular study. Only the results from a good number of replication studies would enable one to begin to say with some confidence how effective research ethics education might be. It should also be kept in mind that studies involving medical ethics (Self et al., 1989) and reviews of broader areas of moral judgment (Leming, 1981; Rest, 1988; Schlaefli et al., 1985) indicate that improving moral judgment through educational interventions is possible, so a no-difference finding in a study using a correlational design, as the present one did, should not be overinterpreted.

Second, the present study--as detailed above under "Limitations"--was far from perfect, mostly owing to the realities of non- (or quasi-) experimental research in real-world settings. In particular, the questionable strength of the "treatment" variable (i.e., the possibly low actual difference between the ethics education experience of students from departments defined as high- versus low-emphasis) mitigates against drawing strong, definitive conclusions from these findings.[10]

A third argument against using the present findings to justify departments' dropping or reducing emphasis on research ethics education grows out of considerations of the relative values of efforts and consequences, and of the limitations of reaching

---

[10]As Shaver (1983) and others have pointed out, verification of independent variable implementation can be a major problem even in experimental research, if the actual implementation is not under the direct control of the experimenter.

conclusions based solely on averages. The situation here could be considered analogous to the way airlines handle instruction in how to use seat belts. Before each flight takes off, airline employees go to great lengths to instruct every passenger how to fasten his or her seatbelt, even though very, very few do not already know how. Airlines do this because the consequences of even a single passenger not knowing how to fasten a seatbelt are potentially very severe, both to the passenger and to the airline--even to the entire industry. Similarly, even if the vast majority of graduate students learn perfectly well the norms of scientific research ethics informally through day-to-day work with their mentors, the consequences of even a few students not learning such norms because that method was inadequate in their particular case, either through their own failings or those of the mentors, are severe enough to justify some form of required formal education in research ethics for all graduate students.

Finally, in considering the implications of the findings of the present study for answering the question of whether they argue for graduate programs reducing emphasis on formal research ethics education, it should be remembered that the main purpose of the present study was to measure mean differences between groups, rather than to assess standards on some absolute scale. Likert-type scaling, as used on the graduate student survey, is well adapted to the purpose of measuring mean differences, and thus to detecting effects of interventions. In fact, a comparative study of Likert, Guttman, semantic differential, Thurstone, and self-rating type scales also demonstrated the Likert scale's superiority at predicting at least one type (voting) of actual behavior (Tittle & Hill, 1967). Nevertheless, Likert-type scales do not necessarily give much information about beliefs in absolute, as opposed to comparative, terms. It follows, then, that one cannot infer from the present study's finding of no difference between students from high- versus low-emphasis departments that students in low-emphasis departments have adequately strict ethical standards, and that therefore departments are justified in confining their efforts at research ethics education to those methods defined in the present study as indicating low emphasis. Logically, one would be just as justified (or just as unjustified) in concluding that the results demonstrate that high-emphasis departments need to redouble their efforts. In fact, the no-difference findings, in themselves, imply neither conclusion.

As for the absolute levels of the ratings themselves, it is beyond the method and design of the present study to say what they do mean. Students from both high- and low-emphasis mechanical engineering departments, for example, gave an average seriousness rating of 5.0 on a 1-to-7 scale. Is that "high enough" to indicate an adequate level of ethical strictness, to indicate that students have adequately learned the ethical norms of scientific research? This is meant as a rhetorical question, and the present author can offer no answer; however, department heads and other administrators responsible for graduate education in the sciences may wish to examine carefully the item-by-item means (Tables C-7 to C-14 in Appendix C) to see if, in their judgment, seriousness ratings of some specific acts may seem lower than they would consider ideal. If so, they may wish to see that their own departments place more emphasis on those particular issues.

Future Studies

As is common in social sciences research, the number of questions and issues raised by the findings of the present study is greater than the number of answers provided by it. Some of these questions and issues stem directly from some unavoidable limitations of the present study. To avoid duplication, these, and some future studies that might address them, are discussed in the section above headed

"Limitations." The present section, in contrast, discusses issues and suggests future studies that arise from the present study's findings, rather than its limitations.

One question immediately arises from the finding that mechanical engineering, as a discipline, appears to place relatively little emphasis on formal education of graduate students in research ethics: Why? It has been suggested to the present author (B. R. Burnham, personal communication, November 27, 1996) that perhaps mechanical engineering, historically, is more of a vocationally oriented field, emphasizing practice, than a research-oriented one, and that therefore it should not be surprising that issues involving research, including research ethics, receive less emphasis in graduate education. Perhaps this is true, but contravening this explanation is the fact that in the present study, 94% of the mechanical engineering respondents indeed reported planning research careers, a virtually identical percentage to those from respondents in physiology and psychology. In fact, one assumes that these students' purpose for earning a Ph.D., traditionally a research degree, is precisely that--to prepare for a research career.

Perhaps the answer is slightly different, residing not in present orientation but rather historical tradition: that mechanical engineering has both a shorter history as a research field than do physiology or psychology, and a shorter history as a research field than as a more applied trade. The present author makes no claim to be an historian of science, and thus freely admits that this hypothesis may have no merit, but notes that history-of-science texts tend to give little mention to engineering per se (e.g., Brush, 1988; Taton, 1966); conversely, however, it should be noted that Hellemans and Bunch (1988) cite Leupold's nine-volume General Theory of Mechanics as the "first systematic treatment of mechanical engineering" (p. 118)--and it dates to 1723. Moreover, it is clear that research in mechanical engineering is grounded in the methodological traditions and even the subject matter of such ancient sciences as physics (especially the mechanics branch), chemistry, and mathematics.

In any case, however, the question of why mechanical engineering appears to place less emphasis on research ethics education is one that should be explored, perhaps through studies involving interviews with faculty, department heads, deans, and officers of accreditation organizations and professional societies (i.e., people who may be expected to be knowledgable concerning the historical traditions of the field). Future research might also profitably study other disciplines that occupy the same "hard-applied" quadrant of Biglan's (1973) typology, especially other branches of engineering, to see if the present study's findings generalize to those disciplines.

Another issue raised by the findings from the department head survey is the need for better, more in-depth measures of the nature of research ethics education in graduate schools. The survey used in the present study was a good first step, but future, larger studies might be able to do more--perhaps soliciting copies of course syllabi, lecture notes, written policy statements, and other such artifacts; interviewing students about the details of their ethics training; determining the actual number of hours devoted to research ethics education in various programs; and so on. Such studies would be quite expensive, both in time and money, but would also give a much more detailed picture of graduate research ethics education than was possible to get from the present study. This more detailed information concerning the true nature of research ethics education in different specific graduate programs would in turn naturally enable better, more sensitive studies of the effectiveness of such education.

Future studies are also suggested by the results from the graduate student survey. For example, the fairly substantial (-0.31) negative effect size for high emphasis in physiology is puzzling. Why should students in departments that give relatively high emphasis to education in research ethics have, on the average, less strict beliefs concerning various acts of possible misconduct? Of course, strictly speaking, this has not been shown--the effect size was not statistically significant and thus, in one sense, might just as well have been zero. It should also be noted that the high-emphasis group included two outliers--respondents whose scores were more than three standard deviations below the mean. Removal of these two (of 65) scores changes the effect size from -0.31 to -0.15--still negative, but much closer to zero. Nevertheless, the original effect size, along with the (statistically significant) effect size of -0.46 on the methodology subscale of the survey, seem large enough to warrant further investigation if this finding proves replicable.

Finally, it may be of interest to explore whether education might possibly affect ethical standards in ways other than the unidimensional one of less versus more strictness of standards. For example, it may be that the norms of science are such that in practice some are considered relatively unimportant, while others are vital (as both common sense and the actual spread in item means in the present study would suggest). Perhaps learning these norms involves learning to make such discriminations, rather than simply moving one's standards to a different position on a unidimensional scale from less strict to more strict. One way to study this might be to gather ratings--using the survey used in the present study--from senior scientists and other presumed experts in scientific research ethics, and see if students who are more experienced, or have more formal ethics training, or both, learn to more closely approximate the distinctions made by their seniors.

Final Summation

To sum up: The present study used a well constructed, designed, and implemented mail survey, and had two main findings. First, the majority of departments in two selected disciplines, physiology and psychology, require some form of formal research ethics education of their Ph. D. students; by contrast, only a very small percentage of mechanical engineering departments requires such formal training. Second, the present study found no evidence that education of Ph. D. students in research ethics has any effect on the strictness of their ethical standards. Departments should not, therefore, simply assume their educational efforts concerning research ethics are effective.

Any single study is inherently limited, including the present one, however diligent were the efforts to carry it out well. There is no substitute for replication; if graduate departments would individually strive to rigorously evaluate their efforts in ethics education, and would make public the results, questions concerning the efficacy of research ethics education could be answered with much more confidence than is possible from any single study.

REFERENCES

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. Psychological Bulletin, 84, 898-918.

Altman, D. (1980). Statistics and ethics in medical research. Misuse of statistics is unethical. British Medical Journal, 28, 1182-1184.

American Psychological Association. (1995). Graduate study in psychology. Washington, DC: Author.

American Statistical Association, Ad Hoc Committee on Professional Ethics. (1983). Ethical guidelines for statistical practice: Report of the Ad Hoc Committee on Professional Ethics. American Statistician, 37, 5-6.

Anderson, M. S., Louis, K. S., & Earle, J. (1994). Disciplinary and departmental effects on observations of faculty and graduate student misconduct. Journal of Higher Education, 65, 331-350.

Aronson, L. (1975). The case of the Case of the Midwife Toad. Behavior Genetics, 5 (2), 115-125.

Bailar, J. C. (1986). Science, statistics, and deception. Annals of Internal Medicine, 104, 259-260.

Barber, B. (1952). Science and the social order. New York: Free Press.

Barber, B., Lally, J. J., Makarushka, J. L., & Sullivan, D. (1973). Research on human subjects: Problems of social control in medical experimentation. New York: Russell Sage Foundation.

Barber, T. X. (1976). Pitfalls in human research. New York: Pergamon.

Barrett, P. M. (1985, May). The premed machine. Washington Monthly, p. 41-51.

Beardsley, T. M. (1988). Truth or consequences? Scientific American, 259 (2), 24.

Bechtel, H. K., & Pearson, W., Jr. (1985). Deviant scientists and scientific deviance. Deviant Behavior, 6, 237-252.

Bell, R. (1992). Impure science: Fraud, compromise and political influence in scientific research. New York: Wiley.

Ben-David, J. (1991). Scientific growth: Essays on the social organization and ethos of science. (Ed. by G. Freudenthal). Berkeley: University of California Press.

Biglan, A. J. (1973). The characteristics of subject matter in different academic areas. Journal of Applied Psychology, 57, 195-203.

Birch, A. J. (1990). Deceit in science: Does it really matter? Interdisciplinary Science Reviews, 15, 334-343.

Blakely, E., Poling, A., & Cross, J. (1986). Fraud, fakery, and fudging. In A. Poling & R. W. Fuqua (Eds.), Research methods in applied behavior analysis (pp. 313-330). New York: Plenum.

Blalock, H. M., Jr. (1987). Some general goals in teaching statistics. Teaching Statistics, 15, 164-172.

Bobys, R. S. (1983). Research fraud factors and effects. Free Inquiry in Creative Sociology, 11, 44-48.

Braunwald, E. (1987). On analysing scientific fraud. Nature, 325, 215-216.

Braunwald, E. (1992). Cardiology: The John Darsee experience. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 55-79). New York: Wiley.

Broad, W., & Wade, N. (1982). Betrayers of the truth: Fraud and deceit in the halls of science. New York: Simon and Schuster.

Brush, S. G. (1988). The history of modern science. Ames: Iowa State University Press.

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand-McNally.

Canary, D. J., & Siebold, D. R. (1984). Attitudes and behavior. New York: Praeger.

Casti, J. L. (1989). Paradigms lost: Images of man in the mirror of science. New York: Morrow.

Chalmers, I. (1990). Underreporting research is scientific misconduct. Journal of the American Medical Association, 263, 1405-1408.

Cohen, R. L., & Ciocca, A. J. (1992). The institutional review board: Ethical gatekeeper. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 204-221). New York: Wiley.

Cournand, A., & Zuckerman, H. (1970). The code of science. Studium Generale, 23, 941-962.

Craig, J. R., & Reese, S. C. (1973). Retention of raw data: A problem revisited. American Psychologist, 28, 723.

Department of Health and Human Services, United States Public Health Service (1989, August 8). Responsibilities of awardee and and applicant institutions for dealing with and reporting possible misconduct in science: Final rule, 42 CFR part 50, subpart A. Federal Register, 54, 32446-32451.

DuBois, B. L. (1989). Accepted practices? A view from outside. Perspectives in Biology and Medicine, 32, 605-612.

Ellenberg, J. H. (1983). Ethical guidelines for statistical practice: A historical perspective. American Statistician, 37, 1-4.

Engler, R. L., Covell, J. W., Friedman, P. J., Kitcher, P. S., & Peters, R. M. (1987). Misrepresentation and responsibility in medical research. The New England Journal of Medicine, 317, 1383-1389.

Faber, B. (1974). The Sloan-Kettering affair [Letter to the editor]. Science, 185, 734.

Fox, M. F. (1989). Disciplinary fragmentation, peer review, and the publication process. The American Sociologist, 20, 188-191.

Fox, M. F. (1990). Fraud, ethics, and the disciplinary contexts of science and scholarship. The American Sociologist, 21, 67-71.

Fox, M. F. (1994). Scientific misconduct and editorial and peer review processes. Journal of Higher Education, 65, 298-309.

Freedman, D. X. (1992). Editorial processes, safeguards, and remedies. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 182-203). New York: Wiley.

Garfield, E. (1980). The ethics of scientific publication. In E. Garfield (Ed.), Essays of an information scientist: Vol. 3. 1977-1978 (pp. 644-651). Philadelphia: ISI Press.

Garfield, E. (1987, April 6). What do we know about fraud and other forms of intellectual dishonesty in science? Part 1. The spectrum of deviant behavior in science. Current Contents, pp. 3-7.

Ghiselin, M. T. (1989). Intellectual compromise. New York: Paragon.

Glass, G. V, & Hopkins, K. D. (1984). Statistical methods in education and psychology (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Goldberger, M. L., Maher, B. A., & Flattau, P. E. (Eds.). (1995). Research-doctorate programs in the United States. Washington, DC: National Academy Press.

Goodstein, D. (1995). The fading myth of the noble scientist. In N. J. Pallone & J. J. Hennessy (Eds.), Fraud and fallible judgment (pp. 21-34). New Brunswick, NJ: Transaction.

Gough, H. G., & Hall, W. B. (1977). A comparison of physicians who did or did not respond to a postal questionnaire. Journal of Applied Psychology, 62, 777-780.

Hagstrom, W. (1965). The scientific community. New York: Basic Books.

Hagstrom, W. (1974). Competition in science. American Sociological Review, 39, 1-18.

Harrobin, D. F. (1989). The grants game. Nature, 339, 654.

Hellemans, A., & Bunch, B. (1988). The timetables of science. New York: Simon and Schuster.

Hersen, M., & Miller, D. J. (1992). Future directions: A modest proposal. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 225-243). New York: Wiley.

Hilgartner, S. (1990). Research fraud, misconduct, and the IRB. IRB: A Review of Human Subjects Research, 12, 1-4.

Huth, E. J. (1986). Irresponsible authorship and wasteful publication. Annals of Internal Medicine, 104, 257-259.

Institute of Medicine (1989). The responsible conduct of research in the health sciences. Washington, DC: National Academy Press.

Kalichman, M. W., & Friedman, P. J. (1992). A pilot study of biomedical trainees' perceptions concerning research ethics. Academic Medicine, 67, 769-775.

Kamin, L. J. (1974). The science and politics of IQ. Potomac, MD: Erlbaum.

Kevles, D. J. (1996, May 27). The assault on David Baltimore. The New Yorker, p. 94-109.

Knight, J. A. (1984). Exploring the compromise of ethical principles in science. Perspectives in Biology and Medicine, 27, 432-442.

Kohlberg, L. (1978). The cognitive-developmental approach to moral education. In P. Scharf (Ed.), Readings in moral education (pp. 36-51). New York: Wiley.

Kohn, A. (1986). False prophets. New York: Blackwell.

Koshland, D. E. (1987). Fraud in science. Science, 235, 141.

Kromrey, J. D. (1993). Ethics in data analysis. Educational Researcher, 22, 24-27.

Kubie, L. S. (1954). Some unsolved problems of the scientific career. American Scientist, 42, 104-112.

LaFollette, M. C. (1992). Stealing into print: Fraud, plagiarism, and misconduct in scientific publishing. Berkeley: University of California Press.

LaPiere, R. T. (1934). Attitudes vs. actions. Social Forces, 14, 230-237.

Leming, J. S. (1981). Curricular effectiveness in moral/value education: A review of research. Journal of Moral Education, 10, 147-164.

Luria, S. (1975, May). What makes a scientist cheat? Prism, p. 15-18.

Mahoney, M. J. (1976). Scientist as subject: The psychological imperative. Cambridge, MA: Ballinger.

Mahoney, M. J., & Kimper, T. P. (1976). From ethics to logic: A survey of scientists. In M. J. Mahoney (Ed.), Scientist as subject: The psychological imperative (pp. 187-193). Cambridge, MA: Ballinger.

Medawar, P. (1990a). The strange case of the spotted mice. In D. Pyke (Ed.), The threat and the glory: Reflections on science and scientists (pp. 71-82). New York: Harper Collins. (Reprinted from New York Review of Books, 15 April 1976, p. 6-11)

Medawar, P. (1990b). Scientific fraud. In D. Pyke (Ed.), The threat and the glory: Reflections on science and scientists (pp. 64-70). New York: Harper Collins. (Reprinted from London Review of Books, 17-30 November 1983)

Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. American Sociological Review, 22, 635-659.

Merton, R. K. (1968). Social theory and social structure. New York: Free Press.

Merton, R. K. (1973). The normative structure of science. In N. W. Storer (Ed.), The sociology of science: Theoretical and empirical investigations (pp. 267-278). Chicago: University of Chicago Press. (Originally published as "Science and technology in a democratic order," Journal of Legal and Political Sociology, 1942, 1, 115-126)

Miers, M. L. (1985). Current NIH perspectives on misconduct in science. American Psychologist, 40, 831-835.

Miller, D. J. (1992a). Personality factors in scientific fraud and misconduct. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 125-139). New York: Wiley.

Miller, D. J. (1992b). Plagiarism: The case of Elias A. K. Alsabti. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 80-96). New York: Wiley.

Miller, D. J., & Hersen, M. (1992). Misconduct and fraud in the empirical sciences: History and overview. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 3-16). New York: Wiley.

Mishkin, B. (1988). Responding to scientific misconduct: Due process and prevention. Journal of the American Medical Association, 260, 1932-1936.

Mitroff, I. I. (1974). Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. American Sociological Review, 39, 579-595.

Mohr, L. B. (1992). Impact analysis for program evaluation. Newbury Park, CA: Sage.

Mohr, M. (1979). The ethics of science. International Science Reviews, 4, 45-53.

National Academy of Sciences, Committee on the Conduct of Science (1989). On being a scientist. Washington, DC: National Academy Press.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine (1992). Responsible science: Ensuring the integrity of the research process (Vol I)., Committee on Science, Engineering, and Public Policy, Panel on Scientific Responsibility and the Conduct of Research. Washington, DC: National Academy Press.

Norusis, M. J. (1993). SPSS for Windows Base System user's guide, release 6.0. Chicago: SPSS Inc.

Petersdorf, R. G. (1986). The pathogenesis of fraud in medical science. Annals of Internal Medicine, 104, 252-254.

Petersdorf, R. G. (1989). A matter of integrity. Academic Medicine, 64, 119-123.

Peterson's Guides, Inc. (1995a). Peterson's guide to graduate programs in the biological and agricultural sciences 1995. Princeton, NJ: Author.

Peterson's Guides, Inc. (1995b). Peterson's guide to graduate programs in engineering and applied sciences 1995. Princeton, NJ: Author.

Poling, A. (1992). The consequences of fraud. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 140-157). New York: Wiley.

Price, A. R. (1994). Definitions and boundaries of research misconduct. Journal of Higher Education, 65, 268-297.

Public Health Service (1986, July 16). Interim policies and procedures for dealing with possible misconduct in science. NIH Guide to Grants and Contracts, 15, 1-37.

Relman, A. S. (1989, April). Fraud in science: Causes and remedies. Scientific American, 260, 126.

Rensberger, B. (1977, January 23). Fraud in research is a rising problem in science. New York Times, pp. 1, 44.

Rest, J. R. (1988, Winter). Can ethics be taught in professional schools? The psychological research. Easier Said Than Done, 22-26.

Robinson, G. M., & Moulton, J. (1985). Ethical problems in higher education. Englewood Cliffs, NJ: Prentice-Hall.

Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. Psychological Science, 5 (3), 127-134.

Schaffner, K. F. (1992). Ethics and the nature of empirical science. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 17-33). New York: Wiley.

Schlaefli, A., Rest, J. R., & Thoma, S. J. (1985). Does moral education inprove moral judgment? A meta-analysis of intervention studies using the Defining Issues Test. Review of Educational Research, 55, 319-352.

Self, D. J., Wolinsky, F. D., & Baldwin, D. C. Jr. (1989). The effect of teaching medical ethics on medical students' moral reasoning. Academic Medicine, 64, 755-759.

Shaver, J. P. (1983). The verification of independent variables in teaching methods research. Educational Researcher, 12 (8), 3-9.

SPSS Inc. (1993). SPSS Base System syntax reference guide, release 6.0. Chicago: Author.

Steneck, N. H. (1984). Commentary: The university and research ethics. Science, Technology, and Human Values, 9 (4), 6-15.

Steneck, N. H. (1993). Fostering integrity in science and engineering research. In Panel on Scientific Responsibility and the Conduct of Research, Committee on Science, Engineering, and Public Policy, National Academy of Sciences, National Academy of Engineering, and Institute of Medicine (Eds.), Responsible science: Ensuring the integrity of the research process (Vol. 2, pp. 3-25). Washington, DC: National Academy Press.

Steneck, N. H. (1994). Research universities and scientific misconduct. Journal of Higher Education, 65, 310-330.

Stewart, W. W. , & Feder, N. (1987). The integrity of the scientific literature. Nature, 325, 207-214.

St. James-Roberts, I. (1976a, September 2). Are researchers trustworthy? New Scientist, 71, 481-483.

St. James-Roberts, I. (1976b, November 25). Cheating in science. New Scientist, 71, 466-469.

Storer, N. W. (1973). The normative structure of science. Prefatory note. In N. W. Storer (Ed.), The sociology of science: Theoretical and empirical investigations (pp. 223-227). Chicago: University of Chicago Press.

Swazey, J. P., Anderson, M. S., & Louis, K. S. (1993). Ethical problems in academic research. American Scientist, 81, 542-553.

Szilagyi, D. E. (1984). The elusive target: Truth in scientific reporting. Journal of Vascular Surgery, 1, 243-253.

Tangney, J. P. (1987). Fraud will out--or will it? New Scientist, 115 (1572), 62-63.

Taton, R. (Ed.). (1966). History of science. (A. J. Pomerans, Trans.). New York: Basic.

Thelen, M. H., & DiLorenzo, T. M. (1992). Academic pressures. In D. J. Miller & M. Hersen (Eds.), Research fraud in the behavioral and biomedical sciences (pp. 161-181). New York: Wiley.

Tittle, C. R., & Hill, R. J. (1967). Attitude measurement and prediction of behavior: An evaluation of conditions and measurement techniques. Sociometry, 30, 199-213.

Toren, N. (1980). The new code of scientists. IEEE Transactions on Engineering and Management, EM-27, 78-84.

Weinstein, D. (1979). Fraud in science. Social Science Quarterly, 59, 639-652.

Wolins, L. (1962). Responsibility for raw data. American Psychologist, 17, 657-658.

Woolf, P. (1981). Fraud in science: How much, how serious? Hastings Center Report, 11 (5), 9-14.

Woolf, P. K. (1988). Deception in scientific research. Jurimetrics Journal, 29 (1), 67-95.

Zuckerman, H. (1977). Deviant behavior and social control in science. In E. Sagarin (Ed.), Deviance and social change (pp. 87-138). Beverly Hills, CA: Sage.

APPENDICES

Appendix A

Department Head Survey and Cover Letters

DEPARTMENT HEAD QUESTIONNAIRE--PSYCHOLOGY

In the questions below, the term **"research ethics education" refers to "the means by which the standards and principles of proper and ethical scientific research conduct are conveyed to graduate students."** Please note that **this study is *not* concerned with professional ethics in applied, treatment, or other activities--only research.**

1. What means of ***research ethics education*** is/are used in your department? (CHECK ALL THAT APPLY.)

_____ **Informal learning through students' ongoing contact** with faculty members.

_____ A ***required* class** specifically dedicated to teaching standards of research ethics.

_____ An ***elective* class** specifically dedicated to teaching standards of research ethics.

_____ **Part of *required* class(es)** primarily dedicated to other content areas.

_____ **Part of *elective* class(es)** primarily dedicated to other content areas.

_____ **Written handbooks or other written policy statements** students are expected to read.

_____ **Attendance at *required* seminars, brown bags, or discussion groups** on standards of research ethics.

_____ **Attendance at *elective* seminars, brown bags, or discussion groups** on standards of research ethics.

_____ None--we do not explicitly teach research ethics, as such.

_____ Other, please describe ____
____

PLEASE CONTINUE ON THE OTHER SIDE OF THIS PAGE!!

2. What organizational or administrative level(s) within the university is/are responsible for **the research ethics education of your department's graduate students**? (CHECK ALL THAT APPLY)

_____ The program (if your department has more than one Ph. D. program area)

_____ The department

_____ The college

_____ The graduate school

_____ The university

_____Other (please describe)_________________________________.

Thank you for your participation. Please return the completed survey in the enclosed envelope, **ALONG WITH A LIST OF THE NAMES OF YOUR DEPARTMENT'S PH. D. STUDENTS. Please EXCLUDE OR SOMEHOW MARK the students in CLINICAL OR COUNSELING psychology, so we do not sample them.**

March, 1996

Dear Colleague:

As you know, in recent years misconduct in scientific research has received much attention from the public, the government, and the mass media, as well as from the scientific community itself. Unfortunately, information is currently lacking concerning **what kind of education in research ethics graduate students are receiving**. To gain needed knowledge in this vitally important area, the Department of Psychology at Utah State University is currently engaged in a study of **research ethics education** in the graduate training of research scientists. As part of this study, we are conducting a survey of university department heads in several disciplines, in order to determine the extent and type of research ethics education in American universities. Your department has been selected to be included in our study.

In order for us to make an accurate assessment of the current situation regarding the education of graduate students in research ethics, it is extremely important that we receive as many completed surveys as possible. We would appreciate your helping us in two ways: (a) take 5 minutes or so to answer the questions on the enclosed survey, concerning your department's current practices with respect to education in research ethics; and (b) return it in the enclosed envelope, ***along with a list of the names of the graduate students in your department who are working toward the Ph. D. Please EXCLUDE or somehow MARK students who are in CLINICAL OR COUNSELING psychology from the listing.***

To help you understand why we need graduate students' names, please allow us to describe very briefly the design of our study:

1. Department heads are being surveyed to obtain important descriptive information concerning the current state of research ethics education in graduate study in the sciences.

2. Department heads' responses will be divided into categories, according to the type of research ethics education the departments provide.

3. Student samples will be randomly drawn from each category; students will be sent questionnaires concerning their opinions toward research ethics issues.

4. Student responses will be compared across categories to determine if their opinions differ.

While obviously the nature of our design is such that your responses cannot be anonymous, we assure you that **responses will be kept completely confidential** by the study team, and names will at no time be associated with results. Individual departments will **not** be named in any written or oral reports from the study. If you have any questions about the study, or our procedures to insure confidentiality, please call Dr. Blaine R. Worthen (801-797-1410), or Perry Sailor (801-797-0090).

We will be grateful if you will help us with this project. To show our appreciation, we would be pleased to send you a summary of the results of our study if you would note on your questionnaire that you would like to receive it.

Sincerely,

Blaine R. Worthen
Professor and Chair,
Research and Evaluation Methodology Program
Co-Principal Investigator

Perry Sailor
Research Associate
Principal Investigator

April 12, 1996

Dear Colleague:

A couple of weeks ago, we sent you a brief survey concerning **research ethics education** in your department's Ph. D. program. We know how busy you must be, so if you have already completed and returned the survey, please accept our sincere thanks and excuse this letter. If not, won't you please take just a few minutes to do so today? In order for us to make an accurate assessment of the current situation regarding the research ethics education of Ph. D. students across the entire United States, it is extremely important that we receive a completed survey from each and every department head selected for the study.

Almost all of the department heads who have completed the survey have complied with our request for the names of their department's graduate students, but a few have declined to do so, citing confidentiality concerns. Again, let us give you our assurance that all responses will be kept completely confidential by the study team, and names of individuals, departments, or universities will at no time be associated with results. Names of graduate students provided to us do not even have their universities identified on our internal records, except by code numbers whose associated university names are kept separately from the student names. **It is extremely important to our study that we be able to survey graduate students; however, if, in spite of our assurances, you feel you cannot provide us with graduate students' names, we would still like for you to answer the two questions on the survey and return it to us.**

If you have any questions, or if for some reason you did not receive the original survey, or it got misplaced, please call Perry Sailor at 801-797-0090 or e-mail him at **perrys@cpd2.usu.edu** and another will be mailed to you immediately.

Sincerely,

Blaine R. Worthen
Professor and Chair,
Research and Evaluation Methodology Program
Co-Principal Investigator

Perry Sailor
Research Associate
Principal Investigator

April 29, 1996

Dear Colleague:

In late March we sent you a brief survey concerning **research ethics education** in your department's Ph. D. program; a couple of weeks ago we sent a followup letter. If you have already completed and returned the survey, please accept our sincere thanks and excuse this third mailing. If you have not, several possible reasons occur to us:

- *You just haven't gotten around to it.* We hope this is the case, and we look forward to getting your completed survey soon.

- *The survey form has been misplaced.* We've attached another one.

- *It will take too much time.* One of us is a former department head, so we understand that there are many demands on your time; we have tried to make the survey as quick and easy to complete and return as possible, and have confined it to only two questions.

- *You're concerned about our procedures for maintaining the confidentiality of the graduate students whose names we are asking for.* We assure you that our procedures for handling the names will insure that they are secure and that to anyone but the two of us, they will be just a list of names unconnected to any location. We have done many, many survey studies, and our record for maintaining the privacy of respondents is completely unblemished. We stand behind our reputations and our past work (and of course, we would never dare do anything unethical in a study of research ethics!) As we explained in our initial letter, Phase 2 of our project involves asking randomly selected graduate students their opinions concerning certain research practices which vary in the degree to which they may be considered misconduct, with selection stratified by their departments' methods of research ethics education. **We want students' names only so that (1) we control the random selection process,** which is essential to good methodological practice, **and (2) we can mail the survey directly to them** without the need for anyone at your end to take time to do anything other than to put an envelope in a mailbox. **In any case, let us emphasize again: If, in spite of our assurances of confidentiality, you do not wish to provide us with graduate students' names, we would still like you to answer the two questions on the attached survey and return it to us.**

In order for us to make an accurate assessment of the current situation regarding research ethics education in Ph. D. programs across the entire United States, it is extremely important that we receive a completed survey from each department selected for the study. The validity of our findings is directly related to the number of responses we receive.

As always, if you have any questions or concerns, don't hesitate to call Perry Sailor at 801-797-0090, or send him e-mail at **perrys@cpd2.usu.edu**.

Sincerely,

Blaine R. Worthen
Professor and Chair,
Research and Evaluation Methodology Program
Co-Principal Investigator

Perry Sailor
Research Associate
Principal Investigator

Appendix B

Graduate Student Survey and Cover Letters

GRADUATE STUDENT QUESTIONNAIRE

A. Do you intend to earn a Ph. D. in your field?

____Yes
____No

B. Do you intend to pursue a career in which at least part of your time will be spent in doing research?

____Yes
____No

C. In what year did you begin graduate study in your present department?

In 19_____ (fill in year)

Now we would like to learn your feelings about several **activities in which a research scientist may engage**. The actions represent a wide range of degrees of misconduct, from slight to extremely serious. Some may not be considered misconduct at all. For each activity listed, **please circle the response that best represents your beliefs about whether or not the activity represents misconduct by the researcher**. In signifying your belief concerning the activity, use the following scale:

NOT AT ALL An Act of Misconduct						An EXTREMELY SERIOUS Act of Misconduct
1	2	3	4	5	6	7

For example, if, in your opinion, a listed activity seems to be **not at all** an act of misconduct, circle "1." If, in your opinion, the activity is a **very slight** act of misconduct, circle "2." If it seems to you to be an **extremely serious** act of misconduct, circle "7," and so on.

	CIRCLE ONE						
1. Ad hominem attacks (i.e., criticizing a person instead of his/her work). . . .	1	2	3	4	5	6	7
2. Applying for funding to support work already done.	1	2	3	4	5	6	7
3. Carelessness in conducting experiments, including reading or recording data	1	2	3	4	5	6	7
4. Double checking <u>only</u> results that don't support one's hypothesis.	1	2	3	4	5	6	7
5. Fabrication of data.	1	2	3	4	5	6	7
6. Failure to inform human subjects adequately.	1	2	3	4	5	6	7
7. Failure to disclose weaknesses in data.	1	2	3	4	5	6	7
8. Failure to disclose weaknesses in research design.	1	2	3	4	5	6	7
9. Failure to make raw data available for re-analysis when requested	1	2	3	4	5	6	7
10. Failure to present results that contradict one's previous research.	1	2	3	4	5	6	7

NOT AT ALL An Act of Misconduct						An EXTREMELY SERIOUS Act of Misconduct
1	2	3	4	5	6	7

		CIRCLE ONE						
11.	Failure to report results that do not support one's hypothesis.	1	2	3	4	5	6	7
12.	Failure to publish until follow-up work is complete.	1	2	3	4	5	6	7
13.	False claims or commitments made in grant proposals.	1	2	3	4	5	6	7
14.	Selective deletion of "outlying" data points.	1	2	3	4	5	6	7
15.	Giving only a cursory review to a paper submitted for publication, if it supports one's own theory.	1	2	3	4	5	6	7
16.	Honest error.	1	2	3	4	5	6	7
17.	Incompetent data analysis.	1	2	3	4	5	6	7
18.	Incomplete authorship (i.e., failure to credit someone who deserves coauthorship).	1	2	3	4	5	6	7
19.	Incomplete documentation of research procedures.	1	2	3	4	5	6	7
20.	Intentional efforts to communicate false or misleading findings.	1	2	3	4	5	6	7
21.	Intentional misinterpretation of results.	1	2	3	4	5	6	7
22.	Misrepresentation of another's work in a citation.	1	2	3	4	5	6	7
23.	Inaccurate representation of research or analysis procedures.	1	2	3	4	5	6	7
24.	Misrepresentation of publication status of an article (e.g., claiming it's "in press" when it has been submitted but not accepted)	1	2	3	4	5	6	7
25.	Mistreatment of human or animal subjects.	1	2	3	4	5	6	7
26.	Breaking down the findings from a single piece of research into multiple papers.	1	2	3	4	5	6	7
27.	Neglect or violation of methodological concerns and procedural precautions, (e.g., loosely following experimental protocol).	1	2	3	4	5	6	7
28.	Overlooking colleagues' use of flawed data, questionable interpretations, or other research transgressions.	1	2	3	4	5	6	7
29.	Plagiarism (i.e., claiming another's work as one's own).	1	2	3	4	5	6	7
30.	Repeated publication of essentially the same content.	1	2	3	4	5	6	7
31.	Reporting the *statistical* significance of an effect while ignoring the magnitude of the effect	1	2	3	4	5	6	7

NOT AT ALL An Act of Misconduct						An EXTREMELY SERIOUS Act of Misconduct
1	2	3	4	5	6	7

		CIRCLE ONE						
32.	Retaliation against whistle-blowers.	1	2	3	4	5	6	7
33.	Reviewing others' work unfairly (e.g., to sabotage a rival)	1	2	3	4	5	6	7
34.	Rigging experiments.	1	2	3	4	5	6	7
35.	False charges of plagiarism against others.	1	2	3	4	5	6	7
36.	Statistical testing of post hoc hypotheses (i.e., of hypotheses made <u>after</u> examining results).	1	2	3	4	5	6	7
37.	Using someone's research ideas without credit.	1	2	3	4	5	6	7
38.	Performing research which ultimately has unintended bad consequences (e.g., work with genetically altered microorganisms unleashing a serious epidemic, or behavioral genetics research leading to discrimination).	1	2	3	4	5	6	7
39.	Underacknowledgment (e.g., failure to cite) of intellectual predecessors, rivals, colleagues.	1	2	3	4	5	6	7
40.	Unjustifiable authorship (i. e., listing someone as an author who was not actually involved in doing the research or writing the article)	1	2	3	4	5	6	7
41.	Using a poor research design.	1	2	3	4	5	6	7
42.	Using one's position to exploit or manipulate others.	1	2	3	4	5	6	7
43.	Using university resources for outside consulting work.	1	2	3	4	5	6	7
44.	Violation of privacy or confidentiality norms regarding subjects.	1	2	3	4	5	6	7

45. Below are listed several possible methods by which graduate students may learn about research ethics. Please check any that you personally have experienced in your graduate training. (CHECK ALL THAT APPLY.)

_____ **Informal learning through ongoing contact** with faculty members.
_____ A **course** specifically dedicated to teaching standards of research ethics.
_____ **Part of a course or courses** primarily dedicated to other content areas.
_____ **Written handbooks or other written policy statements.**
_____ **Attendance at seminars, brown bags, or discussion groups** on standards of research ethics.
_____ **None** -- I have had no specific education in research ethics, as such.
_____ Other, please describe:______________________________

THANK YOU!!! Please return this questionnaire in the envelope provided.

April 1, 1996

Dear ________________:

As you know, in recent years misconduct in scientific research has received much attention from the public, the government, and the mass media, as well as from the scientific community itself. Unfortunately, information is currently lacking concerning graduate students' attitudes about research misconduct. To gain needed knowledge in this vitally important area, the Department of Psychology at Utah State University is currently conducting a survey of **graduate students' attitudes about various kinds of research behavior**. Your name has been given to us by your department head, and you have been randomly selected to receive a survey. **There are several thousand graduate students in the U. S. in your field, but less than 1% have been randomly selected for the survey. Your responses will represent those of many graduate students similar to you. Therefore, it is vitally important that you complete the survey.**

We hope you will help us by doing the following two things: (a) take 15 minutes or so to answer the questions on the attached survey, and (b) return it in the enclosed stamped envelope.

You may notice that there is an identification number on the return envelope. This is so we can keep track of which survey recipients have returned their completed surveys, so we will not have to send them a follow-up. However, we assure you that when we receive your completed survey, we will separate your survey form from the envelope. Thus, it will not be possible to associate a particular survey with a specific individual. Also, we will never write or report anything that could identify any individual respondent, department, or institution. Only summaries of responses for groups of students will be made available.

If you have any questions about the study, or our procedures to insure confidentiality, please call Dr. Blaine R. Worthen (801-797-1447), or Perry Sailor (801-797-0090).

Thank you for helping us with this project.

Sincerely,

Blaine R. Worthen
Professor and Chair,
Research and Evaluation Methodology Program

Perry Sailor
Research Associate
Principal Investigator

July 8, 1996

Dear Graduate Student:

A couple of weeks ago, we sent you a brief survey concerning your opinions about various possible acts of **research misconduct**. We know how busy you must be, so if you have already completed and returned the survey, please accept our sincere thanks and excuse this letter. If not, won't you please take just a few minutes to do so today? In order for us to make an accurate assessment of the beliefs of Ph. D. students across the entire United States concerning issues of research misconduct, it is extremely important that we receive a completed survey from each and every graduate student selected for the study.

If you have any questions, or if for some reason you did not receive the original survey, or it got misplaced, please call Perry Sailor at 801-797-0090 or e-mail him at **perrys@cpd2.usu.edu** and another will be mailed to you immediately.

Sincerely,

Blaine R. Worthen
Professor and Chair,
Research and Evaluation Methodology Program
Co-Principal Investigator

Perry Sailor
Research Associate
Principal Investigator

July 22, 1996

Dear Graduate Student:

In mid-June we sent you a survey concerning **your opinions about various possible types of research misconduct**; then a couple of weeks ago we sent a followup letter. If you have already completed and returned the survey, please accept our sincere thanks and excuse this third mailing. If you have not, several possible reasons occur to us:

- *You just haven't gotten around to it.* We hope this is the case, and we look forward to getting your completed survey soon.

- *The survey form has been misplaced.* We've attached another one.

- *It will take too much time.* One of us is a doctoral student and the other once was, so we understand that there are many demands on your time; we have tried to make the survey as quick and easy to complete and return as possible.

- *You're concerned about confidentiality.* We assure you that our procedures for handling the surveys will insure that they are secure and that *no respondent, department, or university will ever be associated with a particular survey.* We have done many, many survey studies, and our record for maintaining the privacy of respondents is completely unblemished. We stand behind our reputations and our past work (and of course, we would never dare do anything unethical in a study of research ethics!)

In order for us to make an accurate assessment of the beliefs of Ph. D. students concerning issues of research misconduct, it is extremely important that we receive a completed survey from each graduate student selected for the study. The validity of our findings is directly related to the number of responses we receive.

As always, if you have any questions or concerns, don't hesitate to call Perry Sailor at 801-797-0090, or send him e-mail at **perrys@cpd2.usu.edu**.

Sincerely,

Blaine R. Worthen
Professor and Chair,
Research and Evaluation Methodology Program

Perry Sailor
Research Associate
Principal Investigator

Appendix C

Additional Tables

Table C-1

Percentage of Departments Using Various Methods of Research Ethics Education, by Program Effectiveness Rating--Mechanical Engineering

Method	Bottom quartile	2nd quartile	3rd quartile	Top quartile	Overall
Informal learning	100	85	100	100	96
Entire required class	0	0	9	0	2
Entire elective class	0	0	18	9	7
Part of required class	0	0	18	0	5
Part of elective class	11	0	9	9	7
Written handbooks or other statements	22	0	46	18	21
Required seminars / discussions	0	0	9	9	5
Elective seminars / discussions	22	8	27	9	16
None	22	23	9	9	16
Other	11	15	0	27	14
Informal learning <u>only</u>	56	62	18	46	46
Informal learning or None	67	77	18	46	52
Some form required	0	0	36	9	11
Class required	0	0	27	0	7

Table C-2

Percentage of Departments Using Various Methods of Research Ethics Education, by Program Effectiveness Rating--Physiology

Method	Bottom quartile	2nd quartile	3rd quartile	Top quartile	Overall
Informal learning	100	87	86	75	86
Entire required class	24	40	50	60	44
Entire elective class	29	13	7	10	15
Part of required class	18	33	7	20	20
Part of elective class	12	0	7	10	8
Written handbooks or other statements	35	47	36	50	42
Required seminars / discussions	18	13	21	15	17
Elective seminars / discussions	12	27	0	20	15
None	6	13	0	5	6
Other	0	7	7	10	6
Informal learning only	18	20	21	5	15
Informal learning or None	18	20	21	5	15
Some form required	41	60	64	75	61
Class required	41	60	57	75	59

Table C-3

Percentage of Departments Using Various Methods of Research Ethics Education, by Program Effectiveness Rating--Psychology

Method	Bottom quartile	2nd quartile	3rd quartile	Top quartile	Overall
Informal learning	91	96	94	93	93
Entire required class	24	22	6	7	16
Entire elective class	0	4	18	20	9
Part of required class	76	57	65	60	65
Part of elective class	33	26	24	27	28
Written handbooks or other statements	48	52	59	53	53
Required seminars / discussions	5	22	12	20	15
Elective seminars / discussions	19	26	24	20	22
None	0	0	0	0	0
Other	5	4	12	13	8
Informal learning only	5	13	0	0	5
Informal learning or None	5	13	0	0	5
Some form required	86	74	71	67	75
Class required	86	70	71	60	72

Table C-4

Percentage of Departments Using Various Methods of Research Ethics Education, by Size--Mechanical Engineering

Method	Bottom quartile	2nd quartile	3rd quartile	Top quartile	Overall
Informal learning	89	100	91	100	96
Entire required class	0	0	0	8	2
Entire elective class	0	0	9	15	7
Part of required class	0	0	18	0	5
Part of elective class	11	0	9	8	7
Written handbooks or other statements	0	18	27	31	21
Required seminars / discussions	11	9	0	0	5
Elective seminars / discussions	22	9	9	23	16
None	33	0	18	15	16
Other	11	27	0	15	14
Informal learning only	44	55	46	39	46
Informal learning or None	67	55	55	39	52
Some form required	11	9	18	8	11
Class required	0	0	18	8	7

Table C-5

Percentage of Departments Using Various Methods of Research Ethics Education, by Size--Physiology

Method	Bottom quartile	2nd quartile	3rd quartile	Top quartile	Overall
Informal learning	88	94	79	83	86
Entire required class	41	47	50	39	44
Entire elective class	12	18	14	17	15
Part of required class	12	24	7	33	20
Part of elective class	12	0	0	17	8
Written handbooks or other statements	29	59	57	50	42
Required seminars / discussions	18	6	21	22	17
Elective seminars / discussions	18	12	7	22	15
None	18	0	0	6	6
Other	0	12	7	6	6
Informal learning <u>only</u>	24	18	7	11	15
Informal learning or None	24	18	7	11	15
Some form required	53	65	57	67	61
Class required	53	65	50	67	60

Table C-6

Percentage of Departments Using Various Methods of Research Ethics Education, by Size--Psychology

Method	Bottom quartile	2nd quartile	3rd quartile	Top quartile	Overall
Informal learning	95	94	95	89	93
Entire required class	11	17	14	22	16
Entire elective class	0	6	10	22	9
Part of required class	63	67	62	67	65
Part of elective class	37	39	19	17	28
Written handbooks or other statements	26	61	67	56	53
Required seminars / discussions	0	33	10	17	15
Elective seminars / discussions	11	28	10	44	22
None	0	0	0	0	0
Other	5	6	14	6	8
Informal learning only	11	0	10	0	5
Informal learning or None	11	0	10	0	5
Some form required	74	72	76	78	75
Class required	74	67	71	78	72

Table C-7

Mean Scores of Students in High- versus Low-Emphasis Departments on the Graduate Student Questionnaire, by Item--Mechanical Engineering

Item No.	High emphasis (N = 65)		Low emphasis (N = 45)		
	Mean	SD	Mean	SD	Effect size
1	4.7	1.9	5.0	1.8	-0.18
2	5.0	1.7	5.4	1.6	-0.27
3	5.0	1.7	4.9	1.7	0.06
4	4.6	1.6	4.5	2.0	0.03
5	6.9	0.6	6.7	0.8	0.20
6	5.9	1.4	5.9	1.6	-0.03
7	4.9	1.3	5.2	1.6	-0.18
8	5.0	1.3	5.1	1.7	-0.04
9	5.2	1.6	5.0	1.9	0.09
10	5.3	1.3	5.1	1.8	0.09
11	5.1	1.4	5.0	1.7	0.09
12	2.0	1.4	2.3	1.6	-0.17
13	5.4	1.5	5.4	1.7	-0.04
14	5.2	1.8	4.9	2.1	0.15
15	4.5	1.4	4.4	1.9	0.06
16	1.5	1.1	1.8	1.6	-0.20
17	3.6	1.5	3.9	1.9	-0.18
18	5.4	1.4	5.4	1.9	0.00
19	3.9	1.3	4.0	1.8	-0.04
20	6.7	0.8	6.3	1.5	0.22
21	6.6	0.8	6.3	1.4	0.17
22	5.3	1.5	5.3	1.5	-0.01
23	5.1	1.3	4.9	1.7	0.11
24	4.8	1.7	5.1	1.7	-0.14
25	6.4	1.2	5.9	1.8	0.26
26	2.8	1.6	2.9	1.6	-0.05
27	4.4	1.6	4.3	1.8	0.07
28	4.6	1.6	4.8	1.7	-0.10
29	6.9	0.3	6.5	1.5	0.27
30	4.8	1.7	5.0	2.0	-0.11
31	3.8	1.6	3.9	1.9	-0.07
32	6.0	1.4	5.7	1.8	0.15
33	6.4	1.1	6.2	1.5	0.17

(table continues)

Item No.	High emphasis (N = 65)		Low emphasis (N = 45)		
	Mean	SD	Mean	SD	Effect size
34	6.4	1.2	6.3	1.4	0.14
35	6.4	1.1	6.4	1.3	-0.04
36	3.2	1.7	3.6	2.0	-0.19
37	6.0	1.2	5.7	1.6	0.17
38	3.6	2.4	3.5	2.3	0.03
39	4.8	1.5	5.0	1.7	-0.13
40	4.9	1.7	5.0	2.0	-0.06
41	3.3	1.6	3.7	1.8	-0.18
42	5.9	1.3	5.7	1.9	0.11
43	4.9	1.9	4.9	1.9	0.03
44	6.1	1.1	5.9	1.8	0.12

Table C-8

Mean Scores of Students in High- versus Low-Emphasis Departments on the Graduate Student Questionnaire, by Item--Physiology

Item No.	High emphasis (N = 65)		Low emphasis (N = 45)		
	Mean	SD	Mean	SD	Effect size
1	4.5	1.9	4.7	1.8	-0.10
2	4.7	1.7	5.1	1.8	-0.21
3	5.0	1.6	5.5	1.4	-0.34
4	5.0	1.4	4.8	1.6	0.12
5	6.9	0.6	6.9	0.6	0.11
6	6.2	1.2	6.2	1.2	0.05
7	5.0	1.4	5.1	1.3	-0.09
8	4.8	1.5	5.1	1.2	-0.23
9	5.8	1.3	6.0	1.0	-0.17
10	5.4	1.4	5.7	0.9	-0.33
11	5.5	1.4	5.4	1.3	0.09
12	2.2	1.5	2.3	1.6	-0.03
13	5.8	1.6	5.7	1.5	0.02
14	5.0	1.8	5.3	1.6	-0.14
15	4.8	1.5	4.9	1.2	-0.08
16	1.4	0.9	1.4	1.0	-0.02
17	3.5	1.6	4.2	1.7	-0.46*
18	5.5	1.4	5.4	1.3	0.06
19	4.3	1.7	4.4	1.7	-0.09
20	6.8	0.8	6.5	1.2	0.22
21	6.7	0.9	6.7	0.8	-0.05
22	5.4	1.5	5.6	1.4	-0.11
23	5.2	1.4	5.4	1.1	-0.14
24	5.0	1.8	5.3	1.4	-0.23
25	6.6	1.2	6.6	0.8	-0.04
26	3.1	1.7	3.5	1.6	-0.23
27	4.1	1.9	4.7	1.5	-0.41
28	4.9	1.5	5.3	1.2	-0.32
29	6.8	1.1	6.9	0.5	-0.22
30	5.0	1.7	5.5	1.3	-0.41
31	3.9	1.6	4.1	1.7	-0.11
32	6.2	1.3	5.9	1.2	0.19
33	6.5	1.0	6.7	0.5	-0.36

(table continues)

	High emphasis (N = 65)		Low emphasis (N = 45)		
	Mean	SD	Mean	SD	Effect size
34	6.7	0.8	6.7	0.6	0.03
35	6.4	1.3	6.4	1.1	0.05
36	3.6	1.9	3.8	2.0	-0.09
37	6.1	1.3	6.0	1.0	0.02
38	3.4	2.2	4.2	2.4	-0.33
39	4.5	1.7	4.9	1.5	-0.25
40	5.3	1.5	5.4	1.1	-0.05
41	3.1	1.8	3.8	1.7	-0.45*
42	6.2	1.3	6.1	1.2	0.07
43	5.3	1.7	5.5	1.4	-0.15
44	6.4	1.2	6.4	1.0	0.06

* $p < .05$

Table C-9

Mean Scores of Students in High- versus Low-Emphasis Departments on the Graduate Student Questionnaire, by Item--Psychology

Item No.	High emphasis (N = 71)		Low emphasis (N = 68)		
	Mean	SD	Mean	SD	Effect size
1	4.6	1.7	4.6	1.6	0.02
2	5.0	1.5	4.5	1.8	0.26
3	5.0	1.5	4.9	1.4	0.04
4	5.0	1.5	4.9	1.4	0.04
5	7.0	0.3	7.0	0.2	-0.06
6	5.9	1.1	5.9	1.2	0.04
7	4.9	1.3	4.8	1.3	0.05
8	4.8	1.2	4.6	1.4	0.10
9	5.6	1.2	5.8	1.3	-0.13
10	5.2	1.4	5.0	1.4	0.14
11	5.0	1.4	5.0	1.6	-0.01
12	2.1	1.3	1.8	1.2	0.21
13	5.9	1.2	5.7	1.3	0.16
14	5.0	1.6	5.2	1.9	-0.11
15	4.8	1.3	4.7	1.5	0.07
16	1.5	0.9	1.4	1.0	0.05
17	3.9	1.5	3.4	1.7	0.31
18	6.0	1.0	5.8	1.0	0.15
19	4.4	1.2	4.4	1.4	0.00
20	6.7	0.6	6.9	0.3	-0.35
21	6.6	0.7	6.7	0.8	-0.09
22	5.3	1.2	5.2	1.5	0.05
23	5.3	1.2	5.3	1.2	-0.01
24	5.0	1.4	4.9	1.6	0.01
25	6.9	0.6	6.6	0.9	0.23
26	2.7	1.7	3.2	1.8	-0.23
27	4.8	1.4	4.8	1.4	0.01
28	5.2	1.3	5.1	1.3	0.10
29	6.8	0.5	6.9	0.4	-0.23
30	4.6	1.7	4.8	1.5	-0.10
31	3.6	1.5	3.4	1.7	0.14
32	6.0	1.5	6.0	1.3	0.06
33	6.5	0.8	6.5	0.9	0.04

(table continues)

Item No.	High emphasis (N = 71)		Low emphasis (N = 68)		
	Mean	SD	Mean	SD	Effect size
34	6.8	0.7	6.7	0.7	0.14
35	6.7	0.7	6.5	0.9	0.14
36	3.9	1.7	3.3	1.7	0.34
37	6.1	1.0	6.2	1.0	-0.08
38	3.9	2.2	3.4	2.2	0.22
39	4.6	1.6	5.0	1.4	-0.33
40	5.1	1.5	5.1	1.3	-0.07
41	3.0	1.4	3.5	1.8	-0.28
42	6.6	0.6	6.3	1.1	0.30*
43	5.0	1.7	4.8	1.6	0.11
44	6.6	0.9	6.5	0.9	0.09

* $p < .05$

Table C-10

Mean Scores of Students on Items of the Graduate Student Questionnaire, by Discipline

Item No.	Mech. engin. ($\underline{N}$ = 110)		Physiology ($\underline{N}$ = 108)		Psychology ($\underline{N}$ = 139)		
	Mean	SD	Mean	SD	Mean	SD	Post hoc
1	4.8	1.9	4.6	1.8	4.6	1.6	
2	5.1	1.7	4.9	1.7	4.7	1.7	
3	5.0	1.7	5.2	1.5	5.0	1.5	
4	4.6	1.8	4.9	1.5	5.0	1.4	
5	6.8	0.7	6.9	0.6	7.0	0.2	Psy > Phys = ME
6	5.9	1.5	6.2	1.2	5.9	1.2	
7	5.0	1.5	5.1	1.4	4.8	1.3	
8	5.0	1.5	4.9	1.4	4.7	1.3	
9	5.1	1.7	5.9	1.2	5.7	1.3	Psy = Phys > ME
10	5.2	1.6	5.6	1.3	5.1	1.4	
11	5.1	1.5	5.4	1.4	5.0	1.5	
12	2.1	1.5	2.2	1.5	2.0	1.3	
13	5.4	1.6	5.8	1.6	5.8	1.2	
14	5.1	1.9	5.1	1.7	5.1	1.7	
15	4.4	1.6	4.8	1.4	4.8	1.4	
16	1.7	1.3	1.4	0.9	1.4	0.9	
17	3.7	1.7	3.8	1.6	3.7	1.6	
18	5.4	1.6	5.4	1.4	5.9	1.0	Psy > Phys = ME
19	4.0	1.5	4.4	1.7	4.4	1.3	
20	6.5	1.1	6.6	1.0	6.8	0.5	Psy > ME
21	6.5	1.1	6.7	0.9	6.7	0.7	
22	5.3	1.5	5.5	1.5	5.2	1.3	
23	5.0	1.5	5.3	1.3	5.3	1.2	
24	4.9	1.7	5.1	1.7	4.9	1.5	
25	6.2	1.5	6.6	1.0	6.8	0.8	Psy = Phys > ME
26	2.9	1.7	3.3	1.7	3.0	1.8	
27	4.3	1.7	4.3	1.7	4.8	1.4	Psy > Phys = ME
28	4.7	1.6	5.1	1.4	5.1	1.3	Psy = Phys > ME
29	6.7	1.0	6.8	0.9	6.9	0.4	
30	4.9	1.8	5.2	1.6	4.7	1.6	Phys > Psy
31	3.8	1.7	4.0	1.6	3.5	1.6	Phys > Psy
32	5.8	1.6	6.1	1.3	6.0	1.4	

(table continues)

	Mech. engin. (N = 110)		Physiology (N = 108)		Psychology (N = 139)		
	Mean	SD	Mean	SD	Mean	SD	Post hoc
33	6.3	1.3	6.6	1.8	6.5	0.9	
34	6.4	1.3	6.7	0.7	6.8	0.7	Psy = Phys > ME
35	6.4	1.2	6.4	1.2	6.6	0.8	
36	3.4	1.8	3.7	2.0	3.6	1.7	
37	5.9	1.4	6.1	1.2	6.1	1.0	
38	3.6	2.3	3.7	2.3	3.6	2.2	
39	4.9	1.5	4.6	1.6	4.8	1.5	
40	4.9	1.8	5.3	1.4	5.1	1.4	
41	3.5	1.7	3.4	1.8	3.2	1.6	
42	5.8	1.6	6.1	1.2	6.5	0.9	Psy > Phys = ME
43	4.9	1.9	5.4	1.6	4.9	1.7	
44	6.0	1.4	6.4	1.1	6.6	0.9	Psy = Phys > ME

Note. The "post hoc" column shows results of a Student-Newman-Keuls test, where the overall F test was statistically significant. All alpha values are .05.

Table C-11

Items on the Graduate Student Questionnaire, Ranked in Descending Order of Mean Seriousness Rating by Students in Mechanical Engineering

Rank	Item	Mean Rating
1.	Fabrication of data	6.8
2.	Plagiarism (i.e., claiming another's work as one's own)	6.8
3.	Intentional efforts to communicate false or misleading findings	6.5
4.	Intentional misinterpretation of results	6.5
5.	False charges of plagiarism against others	6.4
6.	Rigging experiments	6.4
7.	Reviewing others' work unfairly (e.g., to sabotage a rival)	6.3
8.	Mistreatment of human or animal subjects	6.2
9.	Violation of privacy or confidentiality norms regarding subjects	6.0
10.	Failure to inform human subjects adequately	5.9
11.	Using someone's research ideas without credit	5.9
12.	Retaliation against whistle-blowers	5.8
13.	Using one's position to exploit or manipulate others	5.8
14.	Incomplete authorship (i.e., failure to credit someone who deserves coauthorship)	5.5
15.	False claims or commitments made in grant proposals	5.4
16.	Misrepresentation of another's work in a citation	5.3
17.	Failure to present results that contradict one's previous research	5.2
18.	Failure to make raw data available for re-analysis when requested	5.2
19.	Applying for funding to support work already done	5.1
20.	Failure to report results that do not support one's hypothesis	5.1
21.	Selective deletion of "outlying" data points	5.1
22.	Failure to disclose weaknesses in research design	5.0
23.	Failure to disclose weaknesses in data	5.0
24.	Inaccurate representation of research or analysis procedures	5.0
25.	Carelessness in conducting experiments, including reading or reporting data	5.0
26.	Misrepresentation of publication status of an article (e.g., claiming it's "in press" when it has been submitted but not accepted)	5.0
27.	Unjustifiable authorship (i.e., listing someone as an author who was not actually involved in doing the research or writing the article)	4.9
28.	Underacknowledgment (e.g., failure to cite) of intellectual predecessors, rivals, colleagues	4.9
29.	Using university resources for outside consulting work	4.9
30.	Repeated publication of essentially the same content	4.9
31.	Ad hominem attacks (i.e., criticizing a person instead of his/her work)	4.8
32.	Overlooking colleagues' use of flawed data, questionable interpretations, or other research transgressions	4.7
33.	Double checking only results that don't support one's hypothesis	4.6

(table continues)

Rank	Item	Mean Rating
34.	Giving only a cursory review to a paper submitted for publication, if it supports one's own theory	4.4
35.	Neglect or violation of methodological concerns and procedural precautions (e.g., loosely follwing experimental protocol)	4.3
36.	Incomplete documentation of research procedures	4.0
37.	Reporting the *statistical* significance of an effect while ignoring the magnitude of the effect	3.9
38.	Incompetent data analysis	3.7
39.	Performing research which ultimately has unintended bad consequences (e.g., work with genetically altered microorganisms unleashing a serious epidemic, or behavioral genetics research leading to discrimination)	3.6
40.	Using a poor research design	3.5
41. 42.	Statistical testing of post hoc hypotheses (i.e., of hypotheses made after examining results)	3.4
42.	Breaking down the findings from a single piece of research into multiple papers	2.9
43.	Failure to publish until follow-up work is complete	2.1
44.	Honest error	1.7

Table C-12

Items on the Graduate Student Questionnaire, Ranked in Descending Order of Mean Seriousness Rating by Students in Physiology

Rank	Item	Mean Rating
1.	Fabrication of data	6.9
2.	Plagiarism (i.e., claiming another's work as one's own)	6.8
3.	Rigging experiments	6.7
4.	Intentional misinterpretation of results	6.7
5.	Intentional efforts to communicate false or misleading findings	6.7
6.	Reviewing others' work unfairly (e.g., to sabotage a rival)	6.6
7.	Mistreatment of human or animal subjects	6.6
8.	False charges of plagiarism against others	6.4
9.	Violation of privacy or confidentiality norms regarding subjects	6.4
10.	Failure to inform human subjects adequately	6.2
11.	Using one's position to exploit or manipulate others	6.1
12.	Retaliation against whistle-blowers	6.1
13.	Using someone's research ideas without credit	6.1
14.	Failure to make raw data available for re-analysis when requested	5.9
15.	False claims or commitments made in grant proposals	5.8
16.	Failure to present results that contradict one's previous research	5.6
17.	Misrepresentation of another's work in a citation	5.5
18.	Incomplete authorship (i.e., failure to credit someone who deserves coauthorship)	5.4
19.	Failure to report results that do not support one's hypothesis	5.4
20.	Using university resources for outside consulting work	5.4
21.	Unjustifiable authorship (i.e., listing someone as an author who was not actually involved in doing the research or writing the article	5.3
22.	Inaccurate representation of research or analysis procedures	5.3
23.	Repeated publication of essentially the same content	5.2
24.	Carelessness in conducting experiments, including reading or recording data	5.2
25.	Selective deletion of "outlying" data points	5.1
26.	Misrepresentation of publication status of an article (e.g., claiming it's "in press" when it has been submitted but not accepted)	5.1
27.	Overlooking colleagues' use of flawed data, questionable interpretations, or other research transgressions	5.1
28.	Failure to disclose weaknesses in data	5.1
29.	Double checking <u>only</u> results that don't support one's hypothesis	4.9

(table continues)

Rank	Item	Mean Rating
30.	Failure to disclose weaknesses in research design	4.9
31.	Applying for funding to support work already done	4.9
32.	Giving only a cursory review to a paper submitted for publication, if it supports one's own theory	4.8
33.	Underacknowledgment (e.g., failure to cite) of intellectual predecessors, rivals, colleagues	4.6
34.	Ad hominem attacks (i.e., criticizing a person instead of his/her work)	4.6
35.	Incomplete documentation of research procedures	4.4
36.	Neglect or violation of methodological concerns and procedural precautions (e.g., loosely following experimental protocol)	4.3
37.	Reporting the *statistical* significance of an effect while ignoring the magnitude of the effect	4.0
38.	Incompetent data analysis	3.8
39.	Performing research which ultimately has unintended bad consequences (e.g., work with genetically altered microorganisms unleashing a serious epidemic, or behavioral genetics research leading to discrimination)	3.7
40.	Statistical testing of post hoc hypotheses (i.e., of hypotheses made <u>after</u> examining results)	3.7
41.	Using a poor research design	3.4
42.	Breaking down the findings from a single piece of research into multiple papers	3.3
43.	Failure to publish until follow-up work is complete	2.2
44.	Honest error	1.4

(table continues)

Table C-13

Items on the Graduate Student Questionnaire, Ranked in Descending Order of Mean Seriousness Rating by Students in Psychology

Rank	Item	Mean Rating
1.	Fabrication of data	7.0
2.	Plagiarism (i.e., claiming another's work as one's own)	6.9
3.	Intentional efforts to communicate false or misleading findings	6.8
4.	Rigging experiments	6.8
5.	Mistreatment of human or animal subjects	6.8
6.	Intentional misinterpretation of results	6.7
7.	False charges of plagiarism against others	6.6
8.	Violation of privacy or confidentiality norms regarding subjects	6.6
9.	Reviewing others' work unfairly (e.g., to sabotage a rival)	6.5
10.	Using one's position to exploit or manipulate others	6.5
11.	Using someone's research ideas without credit	6.1
12.	Retaliation against whistle-blowers	6.0
13.	Incomplete authorship (i.e., failure to credit someone who deserves coauthorship)	5.9
14.	Failure to inform human subjects adequately	5.9
15.	False claims or commitments made in grant proposals	5.8
16.	Failure to make raw data available for re-analysis when requested	5.7
17.	Inaccurate representation of research or analysis procedures	5.3
18.	Misrepresentation of another's work in a citation	5.2
19.	Failure to present results that contradict one's previous research	5.1
20.	Overlooking colleagues' use of flawed data, questionable interpretations, or other research transgressions	5.1
21.	Selective deletion of "outlying" data points	5.1
22.	Unjustifiable authorship (i.e., listing someone as an author who was not actually involved in doing the research or writing the article)	5.1
23.	Failure to report results that do not support one's hypothesis	5.0
24.	Double checking only results that don't support one's hypothesis	5.0
25.	Carelessness in conducting experiments, including reading or recording data	5.0
26.	Misrepresentation of the publication status of an article (e.g., claiming it's "in press" when it has been submitted but not accepted)	5.0
27.	Using university resources for outside consulting work	4.9
28.	Neglect or violation of methodological concerns and procedural precautions (e.g., loosely following experimental protocol)	4.8

(table continues)

Rank	Item	Mean Rating
29.	Failure to disclose weaknesses in data	4.8
30.	Underacknowledgment (e.g., failure to cite) of intellectual predecessors, rivals, colleagues)	4.8
31.	Giving only a cursory review to a paper submitted for publication, if it supports one's own theory	4.8
32.	Applying for funding to support work already done	4.7
33.	Failure to disclose weaknesses in research design	4.7
34.	Repeated publication of essentially the same content	4.7
35.	Ad hominem attacks (i.e., criticizing a person instead of his/her work)	4.6
36.	Incomplete documentation of research procedures	4.4
37.	Incompetent data analysis	3.7
38.	Statistical testing of post hoc hypotheses	3.6
39.	Performing research which ultimately has bad consequences (e.g., work with genetically altered microorganisms unleashing a serious epidemic, or behavioral genetics research leading to discrimination)	3.6
40.	Reporting the *statistical* significance of an effect while ignoring the magnitude of the effect	3.5
41.	Using a poor research design	3.2
42.	Breaking down the findings from a single piece of research into multiple papers	3.0
43.	Failure to publish until follow-up work is complete	2.0
44.	Honest error	1.4

Table C-14

Items on the Graduate Student Questionnaire, Ranked in Descending Order of Mean Seriousness Rating by Students in All Disciplines Combined

Rank	Item	Mean Rating
1.	Fabrication of data	6.9
2.	Plagiarism (i.e., claiming another's work as one's own)	6.8
3.	Intentional efforts to communicate false or misleading findings	6.7
4.	Rigging experiments	6.6
5.	Intentional misinterpretation of results	6.6
6.	Mistreatment of human or animal subjects	6.5
7.	False charges of plagiarism against others	6.5
8.	Reviewing others' work unfairly (e.g., to sabotage a rival)	6.5
9.	Violation of privacy or confidentiality norms regarding subjects	6.4
10.	Using one's position to exploit or manipulate others	6.2
11.	Using someone's research ideas without credit	6.0
12.	Failure to inform human subjects adequately	6.0
13.	Retaliation against whistle-blowers	6.0
14.	False claims or commitments made in grant proposals	5.7
15.	Incomplete authorship (i.e., failure to credit someone who deserves coauthorship)	5.6
16.	Failure to make raw data available for re-analysis when requested	5.6
17.	Misrepresentation of another's work in a citation	5.3
18.	Failure to present results that contradict one's previous research	5.3
19.	Inaccurate representation of research or analysis procedures	5.2
20.	Failure to report results that do not support one's hypothesis	5.2
21.	Unjustifiable authorship (i.e., listing someone as an author who was not actually involved in doing the research or writing the article)	5.1
22.	Selective deletion of "outlying" data points	5.1
23.	Using university resources for outside consulting work	5.1
24.	Carelessness in conducting experiments, including reading or recording data	5.0
25.	Misrepresentation of publication status of an article (e.g., claiming it's "in press" when it has been submitted but not accepted)	5.0
26.	Overlooking colleagues' use of flawed data, questionable interpretations, or other research transgressions	5.0
27.	Failure to disclose weaknesses in data	5.0
28.	Applying for funding to support work already done	4.9
29.	Repeated publication of essentially the same content	4.9
30.	Failure to disclose weaknesses in research design	4.9
31.	Double checking only results that don't support one's hypothesis	4.8

(table continues)

Rank	Item	Mean Rating
32.	Underacknowledgment (e.g., failure to cite) of intellectual predecessors, rivals, colleagues	4.8
33.	Giving only a cursory review to a paper submitted for publication, if it supports one's own theory	4.7
34.	Ad hominem attacks (i.e., criticizing a person instead of his/her work)	4.7
35.	Neglect or violation of methodological concerns and procedural precautions (e.g., loosely following experimental protocol)	4.5
36.	Incomplete documentation of research procedures	4.3
37.	Reporting the *statistical* significance of an effect while ignoring the magnitude of the effect	3.8
38.	Incompetent data analysis	3.7
39.	Performing research which ultimately has bad consequences (e.g., work with genetically altered microorganisms unleashing a serious epidemic, or behavioral genetics research leading to discrimination)	3.7
40.	Statistical testing of post hoc hypotheses (i.e., of hypotheses made <u>after</u> examining results)	3.6
41.	Using a poor research design	3.3
42.	Breaking down the findings from a single piece of research into multiple papers	3.0
43.	Failure to publish until follow-up work is complete	2.1
44.	Honest error	1.5

VITA

Perry J. Sailor, Ph.D.

846 N. 4th East
Logan, Utah 84321
801-753-6873

Employment

1994-present: Information Specialist, Early Intervention Research Institute, Utah State University, Logan. Manage longitudinal databases from multiyear studies of effects and costs of early intervention with handicapped children; manage database for meta-analysis of home visiting programs, as well as contributing to addition of new studies to database; designed and maintain system for managing data from national survey of home visiting practices; data analyses on all projects as needed.

1990-1994: Research and Evaluation Associate, Western Institute for Research and Evaluation, Logan, Utah. Among other duties, was responsible for all data management and analysis for a contracted evaluation of computer-assisted instruction software, a study which included over 10,000 students in 16 schools across the nation. Also have assisted in site selection, planning, instrument development, on-site data collection, and report writing for that evaluation and several others.

1988-1990: Research Associate, Arkansas Research and Training Center in Vocational Rehabilitation, University of Arkansas, Fayetteville. Performed all evaluation activities on a project which employed job placement specialists to assist clients who had completed vocational rehabilitation programs. Designed and planned the evaluation, managed all data collection, designed and administered surveys and interviews, and did all data analyses.

1982-1988: Evaluation Associate, Office of Research and Evaluation, Austin Independent School District, Austin, Texas. Did all facets of evaluation, including design, data collection, creation, management, and analysis of small and large data sets, and writing reports for general and professional audiences. Designed and managed computerized system to track District's longitudinal dropout rate, a system which was used as a model for the state by the Texas Education Agency. Served as statistical consultant to committee which designed personnel evaluation system on which merit pay awards were based. Also served as designated information resource person for all secondary level administrators in 60,000 student district. Other evaluation projects included middle schools, magnet schools, Chapter 1 (Title I) programs.

Evaluation Projects Directed

Designed Woodruff School Survey of Parental Opinions Concerning Site-Based Decision Making, Logan City School District.

Evaluation of Early Childhood Home Visit Program, Guadalupe Schools, Salt Lake City, Utah.

Professional Positions

Book Review Editor, *Evaluation Practice*, 1995-
Ad hoc reviewer, *Evaluation Practice*, 1995-
Ad hoc reviewer, *Evaluation and Program Planning,* 1996-

Publications, Professional Papers, Etc.

Sailor, P., Worthen, B. R., & Shin, E. (1997). Class level as a possible mediator of the relationship between grades and student ratings of teaching. Assessment & Evaluation in Higher Education, 22 (3) (accepted for publication)

Worthen, B. R., & Sailor, P. (1995). Review of the Strong Interest Inventory. In J. C. Conoley & J. C. Impara (Eds.), The twelfth mental measurements yearbook (pp. 999-1002). Lincoln, NE: Buros Institute of Mental Measurements.

Parthasarathy, A., Sailor, P., & Worthen, B. R. (1995, April). Effects of respondents' socioeconomic status and timing and amount of incentive payment on mailed questionnaire response rates. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Sailor, P. (1994). Do lefties mature late? An examination of the Koufax phenomenon. Baseball Research Journal, 23, 88-91.

Sailor, P. J., & Worthen, B. R. (1994). A compendium of prior studies on year-round education. Chapel Hill: University of North Carolina, School of Education, North Carolina Educational Policy Research Center.

Worthen, B. R., Van Dusen, L. M., & Sailor, P. (1994). A comparative study of the impact of integrated learning systems on students' time-on-task. International Journal of Educational Research, 21 (1), 25-37.

Burnham, B. R., & Sailor, P. (1993). Site-based decision making: Second annual report to the Utah State Office of Education. Logan, UT: Western Institute for Research and Evaluation.

Worthen, B. R., Van Dusen, L., Leopold, G. D., Sailor, P. J., & Moss, V. D. (1992). Two-year comprehensive assessment of Jostens Learning Corporation's Basic Learning Systems Implementation Models. Logan: Western Institute for Research and Evaluation/Utah State University Department of Psychology.

Worthen, B. R., Van Dusen, L. M., Van Mondfrans, A., Ferguson, T. J., Leopold, G. D., Moss, V. D., Sailor, P. J., Williams, D. D., White, K. R., & Allen, E. (1991). Summary of year-one findings of a two-year comprehensive assessment of Jostens Basic Learning Systems Implementation Models. Logan: Western Institute for Research and Evaluation/Utah State University Department of Psychology.

Doss, D. A., & Sailor, P. J. (1987, April). Counting dropouts: It's enough to make you want to quit too! Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Carsrud, K. B., & Sailor, P. J. (1983, April). Classroom racial composition and the achievement of White, Black, and Hispanic students. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Education

1997: Ph.D. in Research and Evaluation Methodology (REM), Department of Psychology, Utah State University, Logan. All requirements completed for degree to be awarded March, 1997.

1994: M. S., Research and Evaluation Methodology, Utah State University, Logan, 1994.

1978: A. B. (cum laude) in Psychology, Washington University, St. Louis, Missouri.

Honors

National Merit Scholar, 1973-77.